UNIVERSITY*of*
TASMANIA

DOCTORAL THESIS

Probabilistic Methods for Land Cover Change Detection

Author:

Willem OLDING

B.Eng (hons)

Supervisor:

Prof. J.C OLIVIER

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in the

School of Engineering

University of Tasmania - 7/11/2019

Declaration of Originality

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

Signed:
Date: 7/11/2019

Authority of Access

This thesis may be made available for loan and limited copying and communication in accordance with the Copyright Act 1968.

Signed:
Date: 7/11/2019

Statement regarding Published Work contained in Thesis

The publishers of the papers comprising Appendix D and E hold the copyright for that content and access to the material should be sought from the respective journals. The remaining non published content of the thesis may be made available for loan and limited copying and communication in accordance with the Copyright Act 1968.

Signed:
Date: 7/11/2019

"Working smart instead of hard sounds like a great idea, until you realise that you are competing with a lot of very smart people that work extremely hard."

Anon.

"Science is what we understand well enough to explain to a computer. Art is everything else we do."

Donald Knuth

"I love deadlines. I like the whooshing sound they make as they fly by."

Douglas Adams

UNIVERSITY OF TASMANIA

Abstract

Faculty of Science, Engineering and Technology

School of Engineering

Doctor of Philosophy

Probabilistic Methods for Land Cover Change Detection

by Willem OLDING

One of the most powerful tools we possess for global scale monitoring of the surface of the earth are spacebourne remote sensing satellites. Every day they capture vast amounts of data at different spatial, spectral and temporal resolutions which must be processed in order to generate meaningful insights. This thesis focuses on the particular problem of detecting if and when a particular region of land cover experiences a change from one type to another. This problem is made difficult by the fact that the majority of land cover on earth, especially vegetated, undergoes natural variations on both annual and inter-annual time scales driven by changes in season and climate. In this thesis we argue that the key to detecting unnatural changes as accurately and rapidly as possible is to do so with respect to a probabilistic model of the natural variations estimated for the particular region of interest. A method each for change detection, change point estimation and online change monitoring are proposed that follow this strategy. These methods are evaluated on reflectance time series from the Moderate Resolution Imaging Spectroradiometer (MODIS) for two change detection problems, detecting unplanned settlement expansion in South Africa and detecting deforestation of protected areas in Australia. In each case the proposed methods are shown to be effective and require little human supervision suggesting that this approach has potential for use in production systems.

Acknowledgements

The author wishes to acknowledge the following for their contributions:

- Prof. JC Olivier for his guidance, support and patience throughout the time I have known him and in the role of primary supervisor
- Dr Brian Salmon for advice and feedback throughout
- Hydro Tasmania for financial support via the *Dr Sergio Giudici Hydro Tasmania Postgraduate Research Scholarship*
- Council for Scientific and Industrial Research (CSIR) South Africa for kindly sharing their data
- Dr Waldo Kleynhans for his collaboration efforts
- The Department of Primary Industries, Parks, Water and Environment (DPIPWE) for their interest and data sharing

Contents

List of Figures

List of Tables

List of Abbreviations

NDVI	**N**ormalized **D**ifference **V**egetation **I**ndex
EVI	**E**nhanced **V**egetation **I**ndex
MODIS	**MOD**erate Resolutiom **I**maging **S**pectroradiometer
AVHRR	**A**dvanced **V**ery High **R**esolution **R**adiometer
TM	**T**hematic **M**apper
EO	**E**arth **O**bservation
SAR	**S**ynthetic **A**perture **R**adar
PDF	**P**robability **D**ensity **F**unction
GMM	**G**aussian **M**ixture **M**odel
DD	**D**etection **D**elay
RLFA	**R**un **L**ength to **F**alse **A**larm
TPR	**T**rue **P**ositive **R**ate
FPR	**F**alse **P**ositive **R**ate
ROC	**R**eceiver **O**perating **C**haracteristic
AUROC	**A**rea **U**nder **R**eceiver **O**perating **C**urve
IID	**I**ndepdent and **I**dentically **D**istributed

Chapter 1

Introduction

1.1 Problem Statement

Land cover change detection is given as one of the primary aims of numerous earth observation satellite missions. Its focus is to combine multiple data acquisitions to locate regions on the earth where the land cover has been altered in a way that is deemed relevant for a particular application. Change detection is typically used in one of three cases. It may be used as the first stage in a processing pipeline, for example when updating land cover maps it can reduce the problem to only reclassifying regions which have changed. It may also be applied to produce useful information in its own right, for example estimating how much of a particular region has experienced change within a given time span. The third case is where land cover is monitored online in order to detect change shortly after it has occurred. In this case the output of a method is used to guide action in response to the change. Each of these cases has a diverse range of applications that include the monitoring of deforestation, human settlements [1]–[7], fire, agriculture and invasive species [8], [9].

Despite its importance, land cover change detection is an area of research that attracts comparatively few studies compared with other areas of remote sensing such as land cover classification. Historically, due to the limitations of sensors and data processing abilities, land cover change detection methods compared only a few data acquisitions in order to detect change. More recent methods consider many observations of a region as a time series and apply methods from the fields of signal processing and time series analysis.

While many published methods show excellent performance on particular problems none have yet reached the 'holy grail' of land cover change detection: total automation

and high accuracy [10]. While it is unlikely that goal will ever truly be reached there are several characteristics that an ideal method should have. It should be unsupervised, that is not require labeled examples of land cover types or changes, and have a minimum number of parameters that must be set in order to achieve high accuracy. High accuracy can be further decomposed in to a low false alarm rate coupled with a high recall (true alarm rate). Methods having these characteristics have a reduced barrier to usage and can be applied by non-experts in a wide variety of applications.

1.2 Aim and Scope

The aim of this study is therefore as follows:

Problem Statement: To develop methods for land cover change detection that are effective and highly automated.

The study is limited to considering only data from coarse spatial resolution (30m - 1km), high temporal resolution (<5 day sampling period), multi-spectral satellite sensors. This is due to several reasons. Firstly, as mentioned prior, these types of sensors and platforms are designed for large scale, long term monitoring of land cover. The high temporal resolution is of key importance, especially to real time applications, and with current technology this cannot be realised without a trade-off in spatial resolution. Multi-spectral sensors are considered as they are designed with the intention of discriminating different types of land cover by their spectral signatures and have been shown to be very effective for this purpose. Furthermore there is a large catalogue of free, high quality data meeting these requirements as well as upcoming missions ensuring data availability into the future.

1.3 Research Contributions

This body of work contributes two primary contributions to the field of land cover change detection. The first is a method for detecting historical land cover change given a remotely sensed time series of interest and a collection of equal length time series from a

similar geographical region. This method operates by estimating a multivariate Gaussian model with dimensionality equal to the number of samples in the time series. This captures the natural seasonal and inter-annual variations for the region. Time series that do not follow the expected trajectory appear as outliers under the joint distribution and can be detected using outlier detection methods. A method for using this model to also detect the time of change is proposed using the correlation matrix of the estimated model. In this method the correlations between samples before and after a candidate change point are artificially set to zero and the fit of the model re-evaluated for all possible change points. It was found that, for the majority of change signals evaluated, the best fit in terms of maximum likelihood occurs when the candidate change point and real change point are close. The time-of-change can therefore be found with a simple linear search. The findings of this study are published in IEEE Geoscience and Remote Sensing Letters [11] (Appendix D).

The second contribution is that a similar model can be successfully applied to detect changes online for continuous monitoring purposes. This is important for applications where land cover change must be detected shortly after it has occurred. In this method a multivariate Gaussian model is estimated using available historical data for a region and used to make predictions of the next sample of the time series of interest. A time series of the normalised differences between the predicted and observed samples is derived. It was found that by applying classical online change detection methods to this derived time series a substantial increase in detection rate, reduction in false alarm rate and reduction in delay to detection is possible compared with applying a similar detection method to other derived signals suggested in the literature. It is shown that the multivariate Gaussian model of a region is a good predictor that can account for seasonality, inter-annual trends and other disturbances to land cover that act uniformly across the region. Subtracting this variation greatly improves the detectability of disturbances to a single pixel, typically associated with a land cover change event. The findings of this study are published in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing [12] (Appendix E).

1.4 Overview of the Study

The first two chapters provide the reader with a background in earth observation remote sensing and the current state of land cover change detection. Chapter 2 provides the reader with sufficient background in the physics and design of multi-spectral sensors and the satellite platforms that enable earth observation. It also covers the various processing stages required to reduce the effects of atmosphere and geometry in order to produce a signal that is low in unwanted noise. Finally it lists some current earth observation systems and assesses their suitability for detecting changes in land cover. Chapter 3 contains an extensive review of land cover change detection methods and their applications. Methods are considered in the context of how well they address the aim of this study.

Chapter 4 discusses some of the short-fallings of methods in the literature and discusses possible approaches to dealing with them. This leads into the proposal of a method designed to advance towards the goals of automation and accuracy. Chapter 5 highlights the different approach required for change detection in an online setting and gives an extension to the proposed method to allow it to work online. Chapter 6 introduces two change detection problems and gives the results of the proposed methods as well as several other methods from the literature. The final two chapters discuss the results and provide concluding remarks. Chapter 7 gives a detailed analysis and discussion of the results in Chapter 6 and considers how the aims of this study have been addressed. Chapter 8 draws the final conclusions and ideas for the future.

Chapter 2

Satellite Remote Sensing Systems

2.1 Introduction

2.2 Earth Observation Satellite Sensors

Earth observation satellites are by far the most capable tools we possess to remotely observe, measure and quantify large scale changes in our environment. Observing change requires multiple observations that are consistent in location, as well as any other factors not relating to the phenomenon being measured. There are several properties of satellite systems that make them so well suited to this task:

Continuity - Once in orbit satellite platforms can maintain operations for sometimes decades at a time. These long term data sets are crucial for monitoring long term trends. Furthermore new satellite platforms are often designed such that some of their products can merge seamlessly with older missions providing inter-mission continuity. Examples of this are the Landsat missions 1-8, which span over 40 years, and the MODIS -> VIIRs continuity mission [13], [14].

Consistency - Satellite systems are capable of maintaining an almost constant orbit allowing locations to be repeatedly measured from the same altitude and in many cases identical viewing angle through time. Keeping these factors constant reduces the corrections that must be applied to each measurement in order to yield a time series where the signal is dominated by the phenomenon of interest. Some factors such as atmospheric conditions and solar irradiation angle cannot be controlled. Section 2.3 discusses various method for correcting for these factors to reduce their impact on the final signal.

Coverage - With a suitable orbit and sensor geometry it is possible for a satellite platform to image the entire earth within a relatively short space of time. These global data sets are archived, catalogued and in many cases made available for free through open access. This significantly reduces the entry barrier for conducting scientific studies over large regions where previously expensive aerial surveys or in-situ data collection would have been required [15].

Space programs from major governments of the world have dedicated earth observation programs for measuring land cover, oceans, atmosphere, polar caps, magnetic fields, elevation and climate. These diverse applications require different types of sensors which can be roughly catagorised into the hierarchy of figure 2.1. At the top level of the hierarchy a sensor can be classified as either *passive* or *active*. Active sensors emit a known signal and measure the incident energy that is reflected back to the sensor by the object being measured. Examples of active sensors are RADAR and LIDAR. Applications of active sensors include weather observation, Digital Elevation Models (DEMs), 3D Urban mapping and forest structure mapping. Passive sensors do not emit energy by themselves but measure incident energy from other sources. Almost always this energy takes the form of electromagnetic radiation which is emitted directly by the object being measured or reflected from another source, which is almost always the sun. Spectroradiometers are a subtype of passive sensor which splits the incident radiation at the sensor into a number of spectral bands. This allows for the identification of objects based on spectral signatures making them particularly suitable for remote sensing of land cover and vegetation. Applications of passive sensors include vegetation mapping, atmospheric sensing, ocean sensing, temperature measurement and geological surveys. This thesis is concerned exclusively with the measurement of vegetated land cover and so will limit our discussions only to the workings of satellite based passive spectroradiometry.

2.3 Physics of Passive Spectroradiometry

Remotely sensed data from spectroradiometers is typically available in multiple forms each with different degrees of pre-processing applied. The most unprocessed form of the data is the raw digital values output by the sensor that vary in response to the incident radiation. These are known as Digital Numbers (DNs) in the remote sensing field. These

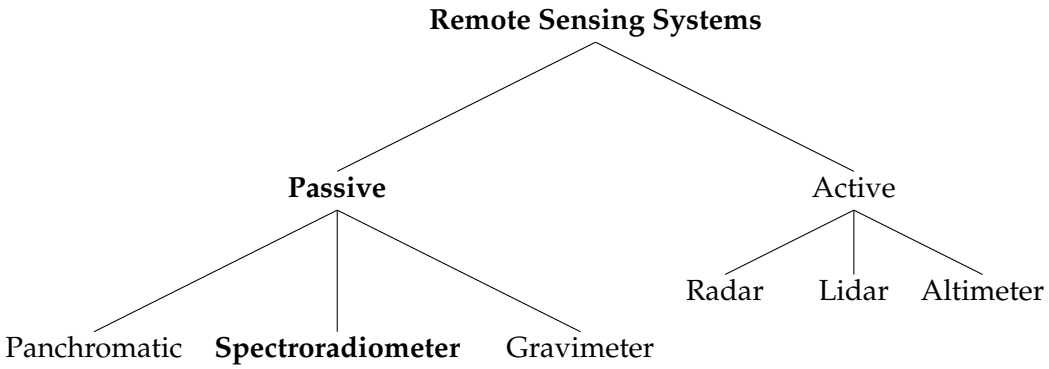


FIGURE 2.1: Classification tree of remote sensing systems.

DNs given in isolation have no physical interpretation. Each level of preprocessing aims to take these DNs closer to a particular physical property of interest. The processing should simultaneously reduce the influence of unwanted factors such as changes in solar geometry, atmospheric conditions and viewing angle. In this section the various levels of data pre-processing are described along with the models of the physical phenomena on which they are based. For further background see the excellent introduction to the physics of remote sensing systems [16].

2.3.1 Digital Numbers and Sensor Radiance

At the most basic level the digital number registered by the sensor can be converted to at-sensor-radiance by normalising with constants found during the pre-launch calibration phase. In many cases the DN response can be assumed to be linearly proportional to *Radiance*. For example from the Landsat 7 data users handbook [17], the at sensor radiance for a given spectral band is calculated by

$$L = \left(\frac{L_{\text{max}} - L_{\text{min}}}{DN_{\text{max}} - DN_{\text{min}}} \right) (DN - DN_{\text{min}}) + L_{\text{min}} \tag{2.1}$$

where DN_{min} and DN_{max} are the minimum and maximum allowable DN values, L_{min} and L_{max} are the calibrated reflectance values of the minimum and maximum DNs for the band of interest and DN is the digital number of the measurement. One thing to consider is that this conversion is only valid for radiance between L_{min} and L_{max}. Most sensors have a saturation flag that is set if the incident radiance is outside the calibrated values. This allows it to be dealt with in later processing stages.

Sensor radiance is a measure of the incident energy per unit time, per unit area, per unit solid angle and has the units Watts per steradian per meter square, $Wsr^{-1}m^{-2}$. To understand how this is derived consider a unit area of surface dA. The energy of the radiation received by this surface per unit time is known as the *Irradiance*, E. It is also sometimes referred to as the radiant flux density. This has the units Watts per meter square, Wm^{-2}.

Irradiance considers all radiation received by the surface irrespective of direction. For photographic and remote sensing applications it is important to also consider the direction of radiation such that the radiation emanating from a single region of the earth can be considered. Consider a range of directions from the surface normal $\theta + d\theta$, and around the surface normal $\phi + d\phi$. These two angles define a solid angle, $d\Omega$, given by

$$d\Omega = \sin\theta d\theta d\phi \tag{2.2}$$

which has units of Steradians (sr).

From these quantities the radiance is given by

$$L = \frac{d\Phi}{dAd\Omega\cos\theta} \tag{2.3}$$

This can also be expressed as a function of wavelength giving the spectral radiance L_λ with units $Wsr^{-1}m^{-3}$.

If the radiance is integrated over the hemisphere of all possible angles this gives the irradiance thus the relationship between radiance and irradiance as

$$E = \int_{\theta=0}^{\frac{\pi}{2}} \int_{\phi=0}^{2\pi} Lcos\theta d\Omega. \tag{2.4}$$

2.3.2 Reflectance

One of the most commonly used measures in passive remove sensing is reflectance. The reflectance is the ratio of reflected radiation to incident radiation of a surface and defined in terms of radiance. Reflectance is particularly useful in remote sensing as it is a property of the object or surface being measured. Reflectance as a function of wavelength (spectral reflectance) is the key measure used to identify objects by radiometric remote sensing.

2.3.2.1 Top-Of-Atmosphere Reflectance

The at-sensor radiance can be converted to top-of-atmosphere reflectance by correcting for differences in illumination from the earth-sun distance, solar zenith angle and solar spectral irradiance. Top-of-atmosphere refers to the fact that no attempt has been made at correcting for atmospheric effects. The distance and angle parameters are a function of time but are simple to calculate due to the regular nature of orbital mechanics. Let $E_{sun,\lambda}$ be the mean solar spectral irradiance at the top of the earth's atmosphere. The top-of-atmosphere reflectance can be calculated as

$$\rho_\lambda = \frac{\pi d^2 L_\lambda}{E_{sun,\lambda} \cos\theta} \tag{2.5}$$

where d is the earth-sun distance in astronomical units, θ is the solar zenith angle. This quantity is sometimes also called effective planetary reflectance.

2.3.2.2 Bottom-Of-Atmosphere Reflectance

Bottom-Of-Atmosphere correction aims to give the reflectance that would be observed if there was no atmosphere present. This is a much more complex correction as it requires measurements of the atmosphere. Often these are provided by sensors onboard the same satellite platform but may also rely on other platforms.

The sun emits electromagnetic radiation mostly at wavelengths between 0.1μm and 1000μm. The spectral curve of sunlight is shown in figure 2.2. Upon reaching the earth the incident radiation must traverse the earth's atmosphere where it experiences a combination of atmospheric absorption and scattering.

When an electromagnetic wave interacts with a molecule of gas under the right conditions the molecule can absorb the wave in order to increase its own energy. It does this be either increasing its electrons energy levels, vibration, rotation or a combination. Quantum mechanical laws state that these increases in energy can only occur in discrete levels which correspond to discrete wavelengths. This leads to characteristic spectral absorption lines corresponding to the different excitations of different molecules in the atmosphere. See [16] Tab. (4.2) for a tabulation of the major absorption lines of the earth's atmosphere due to different gases.

Atmospheric scattering can be caused both by gaseous molecules, microscopic particles and aerosols. The molecular component can be modeled as Rayleigh scattering [16]. In a Rayleigh scattering model the gaseous molecules considered as electrically conducting spheres that are excited by incident EM radiation. They then re-radiate the energy at the same wavelength in all directions. The scattering cross section for a sphere of size d by radiation with wavelength λ is given by

$$\sigma_s = \frac{128\pi^5 d^6}{3\lambda^4} \tag{2.6}$$

assuming an infinite refractive index [16]. This value can be used to calculate the fraction of light scattered per volume The inverse relationship to the fourth power of the wavelength indicates that Rayleigh scattering mostly occurs in shorter wavelengths.

While the molecular composition of the atmosphere is quite uniform across the earth the presence of aerosols and particles if much more spatially heterogeneous and is dependent on local land cover and weather.

Figure 2.3 shows the attenuation of the atmosphere to EM radiation as a function of wavelength. This highlights the regions of the spectrum that are useful for remote sensing of land cover.

The atmospheric correction algorithm applied to generate the MODIS surface reflectance product is described in [18]. It makes use of data from other MODIS products (MOD04 - spectral aerosol optical thickness, MOD05 - precipitable water, MOD07 - ozone), Digital Elevation Models (DEMs) as well as surface measurements of water vapor, ozone and aerosols. The effects of Rayleigh scattering, tropospheric aerosols, gaseous absorption and light cirrus clouds are modeled and the models inverted such that their effects are removed from the final product.

2.3.2.3 Bi-directional Reflectance Distribution Function

The major remaining source of unaccounted for variance is the sensor viewing angle. Scanning sensors, both in-track and across-track, by design do not capture all measurements from the same angle. A larger swathe width allows for the re-imaging of entire earth surface within fewer orbits however subsequent acquisitions of a given area will have different viewing angles and this must be accounted for. Perfect knowledge of the

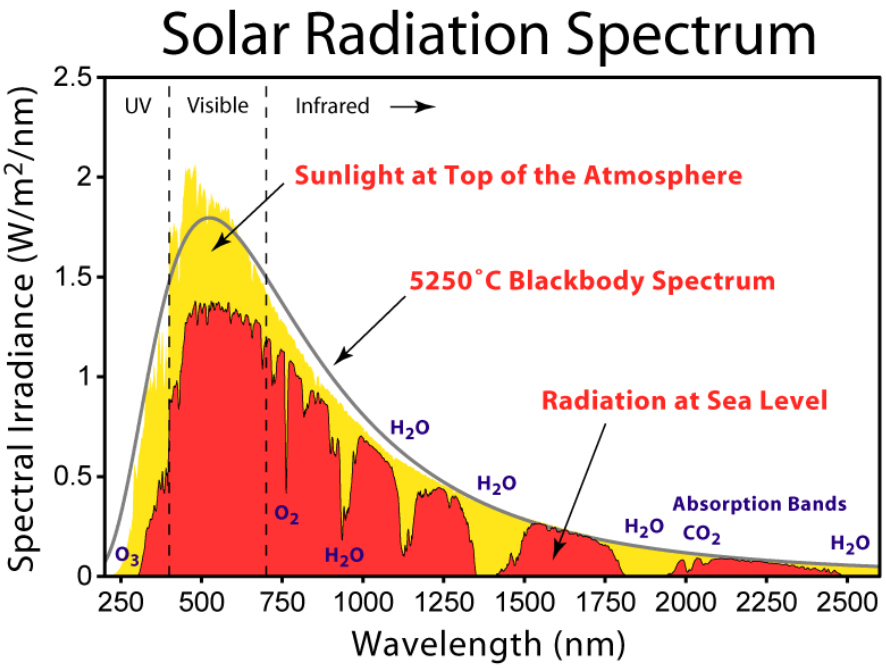


FIGURE 2.2: Spectrum of solar radiation at the top of atmosphere and at the surface of the earth. Adapted from original figure by Robert A. Rohde (https://commons.wikimedia.org/wiki/File:Solar_Spectrum.png) under Creative Commons Attribution 3.0

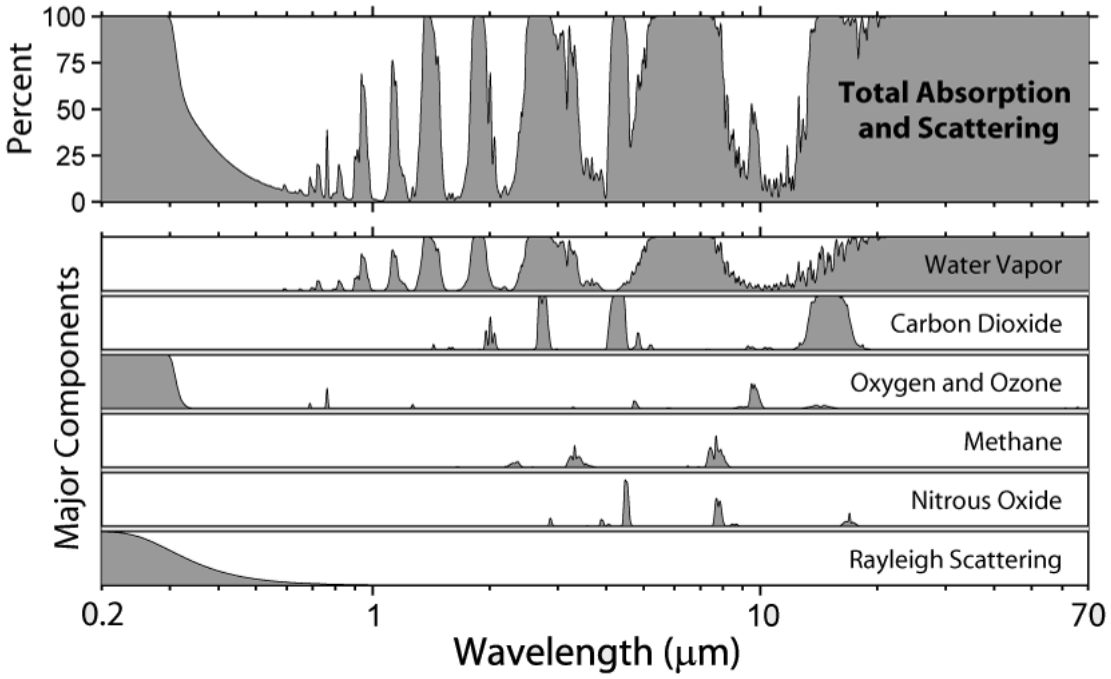


FIGURE 2.3: A breakdown of the contributions different atmospheric gases and Rayleigh scattering to total atmospheric transmission. Adapted from original figure by Robert A. Rohde (https://commons.wikimedia.org/wiki/File:Atmospheric_Transmission.png) under Creative Commons Attribution 3.0

relationship between the angles of incident radiation and measured radiation for a given wavelength allows for converting measurements of reflectance to equivalent measurements at a different observation angle. Ideally all measurements should be corrected to appear as if they were taken from directly above the object, known as nadir view.

Upon encountering a boundary between two media such as the air and earth surface the EM radiation experiences one or both of two phenomena, reflection/transmission and scattering. If the boundary between the two media is sufficiently smooth the interaction can be described by reflection and transmission. Consider incident radiation striking the boundary at some angle θ_i to the boundary normal vector. Reflected radiation will have the angle $\theta_r = -\theta_i$. Radiation that is transmitted through the boundary will be refracted and have an angle given by Snell's Law. Let n_1 and n_2 be the refractive indices of the two media. The transmitted radiation angle is given by

$$\sin\theta_t = \sin\theta_i \frac{n_1}{n_2} \tag{2.7}$$

This type of interaction is important for describing the interfaces between air and water. Most land cover types are not sufficiently smooth to follow this model and can be better described by scattering.

Scattering is the major mode of interaction for describing the interaction of EM radiation and rough surfaces such as land cover. If, as before, we consider an incident beam of radiation at an angle θ_i to the surface normal and with flux density F. In this model the radiation is scattered in a range of angles and the Bi-directional Reflectance Distribution Function (BRDF) describes how the incident irradiance is distributed across the range of angles. Let $E = F\cos\theta_i$ be the irradiance of the surface from the incident mean. Consider also some scattered radiance L_s at some angle θ_s over a solid angle Ω. The BRDF relates these two quantities as a function of the angles

$$f(\theta_i, \theta_s) = \frac{L_s}{E}. \tag{2.8}$$

If the BRDF for a given surface is known exactly it is possible to calculate the scattered radiance at any angle for a given incident irradiance at a specific angle.

The equation above describes a two dimensional BRDF. For the three dimensional

case the BRDF is a function of two additional angles describing the azimuthal direction of the incident and scattered radiation, $f(\theta_i, \phi_i, \theta_s, \phi_s)$. This allows for describing surfaces that have different roughness depending on the direction. Furthermore we expect the BRDF to also be a function of wavelength. If the surface roughness is isotropic the BRDF can be re-parameterised in terms of the azimuth angle between incident and scattered rays.

The perfectly reflective surface has a BRDF of a Dirac delta at $\theta_i = -\theta_s$. Another interesting case is where the scattered radiance is isotropically distributed over the entire hemisphere normal to the surface. This is known as Lambertian Scattering and has a BRDF that is constant. The BRDF of real surfaces ranges between these two extremes.

[19] describes the necessity to perform viewing angle corrections:

> The angular reflectance of the land surface changes not only with viewing position, but also with the position of the source of irradiance in the hemisphere. Thus, both the radiance in a specific direction and the albedo of a surface are dependent on the sun's position in the sky. The angular pattern of downwelling diffuse radiance, which depends largely on the state of the atmosphere, will also influence both the albedo and the specific surface radiance observed in a given direction. This means that the atmospheric state (e.g., turbidity) must be taken into account when BRDFs and the surface albedo measures derived from them are extracted from a series of directional observations.

The key is that perfect knowledge of the BRDF of the land surface allows for converting the reflectance to equivalent reflectance at nadir view. Surface reflectance that has been corrected in this way is referred to as Nadir BRDF Adjusted surface Reflectance (NBAR). Multiple samples of a single patch of land cover from different viewing angles can be used to calculate an approximation to the surface BRDF. An example of estimating a BRDF for correction is given in [20]. This is the approach taken by the MODIS MCD43A4 product to apply viewing angle corrections. A linear weighted sum of kernels is used to approximate the BRDF and the kernel weights are estimated for each measurement by minimizing the error in a least squares sense based on all available measurements in a 16 day temporal acquisition window. The kernel approximation for the

MCD43A4 product is given by

$$f(\theta, v, \phi, \lambda) = \alpha_{\text{iso}}(\lambda) + \alpha_{\text{geo}}(\lambda)K_{\text{geo}}(\theta, v, \phi, \lambda) + \alpha_{\text{vol}}(\lambda)K_{\text{vol}}(\theta, v, \phi, \lambda) \tag{2.9}$$

where θ, v, ϕ are the solar zenith, view zenith and relative azimuth angles and λ is the wavelength. α_{iso} is the isotropic weight of the BRDF, α_{geo} and α_{vol} are the weights for the surface scattering and volume scattering kernels respectively [20]. These kernels are defined based on knowledge of their respective scattering mechanisms. This model is well suited to processing large scale remote sensing data as the correction only requires estimating three parameters per band using a single matrix inversion. The calculated BRDF parameters for each measurement are available in the MCD43A1 product and can themselves be used to derive useful properties of the land cover.

2.3.3 Derived Indices

In addition to using single bands on their own it is commonplace to derive additional signals by combining bands together. These are usually targeted to specific applications and are designed in such a way as to amplify phenomenon of interest while reducing the influence of extraneous factors. Indices have been designed specifically for detecting vegetation, water, fire/burn scars, snow, clouds and more. Only the two most prevalent indices designed for vegetation will be discussed here.

By far the most commonly used of these is the Normalised Difference Vegetation Index (NDVI) [21]. This is defined as the ratio of the difference between reflectance in the Near Infrared (NIR) and red bands divided by their sum. For MODIS this corresponds to bands 2 and 1 respectively giving the NDVI formula as

$$NDVI = \frac{B2 - B1}{B2 + B1}. \tag{2.10}$$

where $B1$ and $B2$ are observations from bands 1 and 2 respectivly for a single pixel at a single time sample.

The normalisation by the sum aims to reduce any residual influence of solar angle and atmosphere while enhancing the sensitivity to the chlorophyll absorption spectral

properties that characterise vegetation [21]. The effectiveness of this index has seen it become ubiquitous in multi-spectral studies of vegetation.

The Enhanced Vegetation Index (EVI) additionally makes use of the blue band (MODIS band 3) to further reduce the influence of the atmosphere, specifically aerosols [22]. The formula for EVI for MODIS is given by

$$EVI = \frac{B2 - B1}{B2 + C_1 B1 - C_2 B3 + L} \tag{2.11}$$

where C_1, C_2 and L are sensor specific constants ($C_1 = 6$, $C_2 = 7.5$ and $L = 1$ for MODIS).

2.4 Earth Observation Missions

At the time of writing there are over 600 earth observation satellites in orbit [23]. The majority of these are for the purpose of optical imaging making them unsuitable for the aims of this thesis. Of the remainder many are for commercial use only.

Throughout this thesis data from the MODerate resolution Imaging Spectroradiometer (MODIS) platform is used exclusively. In this section this platform and several others are discussed. We limit ourselves to four platforms that provide open access, multi-spectral, high temporal resolution data. Arguments are made justifying the choice of data from the MODIS program for change detection.

2.4.1 MODIS

The MODIS program refers to two almost identical instruments onboard the Terra and Aqua satellite platforms. Launched in December 1999 and May 2002, for Terra and Aqua respectively, their main objective is to provide high temporal resolution global products to observe the long term dynamics of land, ocean and atmosphere. The Terra platform operates in a low earth orbit at an altitude of 705km and with an orbit period of 99 minutes. Combined with the rotation of the earth this allows the MODIS sensor to make repeat observations of a single point on the earth with a period between 1-2 days. The Aqua platform has an almost identical orbit and combined they can produce daily observations of any region on earth. MODIS uses a cross-track scanner with a scanning angle of ± 55 degrees. This corresponds to an on ground swathe width of approximately 2330km.

The spatial resolution of the imagery is dependent on the spectral band of interest. The MODIS program also includes services for storage and processing of data for different applications. These MODIS products allow researchers and practitioners to have fast access to identically processed data sets. This study makes use of the MCD43A4 product [24] which contains BRDF corrected surface reflectance data for bands 1-7 at 500m resolution. This product has a temporal resolution of 8 days where the value for each pixel is selected as the highest quality acquisition within a temporal window of 16 days.

2.4.2 VIIRS

The Visible Infrared Imaging Radiometer Suite (VIIRS) is a sensor onboard the Suomi National Polar-orbiting Partnership (NPP) platform and also on the upcoming Joint Polar Satellite System (JPSS). The VIIRS system was designed to provide continuity with MODIS sensors for long term global earth observation. The VIIRS was launched in October 2011 and the data catalogue begins at Jan 2012 [25]. The resulting data is very similar to that generated by MODIS sharing the same spatial resolution and 8-day temporally aggregated, BRDF corrected products. For experiments on time series the lack of historical data compared with MODIS makes it a less desirable choice for this study. It is anticipated that this platform will replace or at least complement MODIS and should be considered for implementing real-time applications.

2.4.3 Landsat Missions

The Landsat program is the longest continuous earth observation mission. While the earliest Landsat platform was launched in 1972 the data continuity missions begin with Thematic Mapper (TM) sensor which was first included on the Landsat 4 platform (1982). While subsequent platforms contain additional advanced sensors they maintain data compatibility with the original thematic mapper. This continues to the present day (as of 2018). While mainly designed for land cover mapping purposes it has found several uses for detecting land cover change, mostly using image pair analysis methods rather than time series [26].

The spatial resolution of the TM sensor is 30 m which gives far greater granularity and like MODIS there is significant infrastructure to pre-process, store and distribute the

data from the Landsat missions. The Landsat 4-8 platforms also occupy a low earth orbit at around 700km altitude however its orbit and swath width is such that a complete re-imaging of the earth takes around 16 days. This makes it much less suitable for applications requiring time series of data, especially considering that a temporal window approach would also be considered in order to reduce the probability of a cloud-free acquisition. The low temporal resolution make Landsat a poor choice for time series based method.

2.4.4 Sentinel Program

The Sentinel program comprises of many different satellite systems for observing different aspects of the earth and atmosphere. The Sentinel-2 mission is comprised of two near-identical platforms in low earth orbit. It is designed for land cover observation and is comparable to the Landsat missions. Sentinel-2A was launched June 2015 and 2B in March 2017 [27] with plans for at least two additional platforms to be added in the future.

The revisit time for Sentinel-2 is 5 days between the two platforms placing it in-between Landsat and MODIS. The inclusion of additional platforms has the potential to reduce this further. For the vegetation sensitive bands the spatial resolution is superior to both at 20m. The Sentinel program is very new and therefore unsuitable for experiments on time series covering multiple years. Much like VIIRS it is anticipated that this will become significant in future studies and should be considered for current implementations of real-time monitoring.

2.5 Justification of use of MODIS

At the time of writing MODIS remains the most suitable source of data for remotely sensed time series for several reasons. Perhaps the most important factor is the large existing catalogue of data spanning from around 2002 to the present day. As will be seen in later chapters this wealth of historical data is what enables estimation of how land cover responds to environmental effects under no-change conditions. This would be much less feasible using VIIRS or Sentinel which provide only a few years worth.

There is also the closely related factor of temporal resolution. While Landsat provides an extensive historical catalogue and has a much higher spatial resolution it only yields approximately 22 acquisitions per year compared to the 360 from MODIS. This limits the temporal resolution with which change events can be located, decreases the probability of obtaining an acquisition free from clouds and leads to an inescapable increase in the delay between the occurrence of a change and its detection for online methods.

Another factor is to allow for fair comparisons with existing work. Of the surveyed studies in the literature that apply reflectance time series for detecting land cover change the vast majority make use of MODIS with only a few exceptions (see Chapter 3). As such it makes sense to control this variable to allow for a fair comparison. The methods developed in this work are by no means MODIS specific and can be adapted to other platforms as more data becomes available.

2.6 Conclusion

This chapter covered the relevant theory required to understand how digital measurements taken by a satellite sensor are corrected for solar geometry, atmospheric effects and viewing angle such that the influence of these unwanted factors on the signal is reduced. This is particularly important for change detection as any variations in the signal that are not a result of changes in the underlying land cover decrease the effective signal-to-noise ratio. Through techniques of modelling atmospheric effects and properties of the surface being measured it is possible to approximate the measurements as being taken from directly above the object in the absence of atmosphere. Provided these corrections are sufficiently effective, signals composed of a number of these corrected observations over time should have a decrease in noise caused by unwanted factors compared with the uncorrected observations. This makes it far better suited for long term change detection applications.

Four candidate sources of data are discussed and evaluated based on factors such as temporal resolution, spatial resolution, spectral resolution and historical data availability. It is concluded that the MODIS platform is by far the most suitable data source at this time for long term change detection. It is anticipated that this will change as the data catalogues for other platforms continue to grow.

Chapter 3

Land Cover Change Detection

3.1 Introduction

It is easy to imagine why land cover change may be hard do define. The surface of the earth can be highly dynamic. Deciduous forests lose their leaves, mountain tops can change from rocky to snow covered and crops are periodically sown and harvested. Are these examples of land cover change or part of a periodic cycle? Land cover is constantly changing therefore the definition of a detection problem requires specifying which types of change are to be detected, and which types are to be ignored.

Each of the studies reviewed in this chapter are considered in the context of how they are advancing the field toward the goals of total automation and high accuracy. Automation is discussed in terms of the level of supervision the method requires as well as the number of parameters that must be selected to achieve high accuracy. The accuracy aspect is considered as two separate but equally important goals: How the changes of interest are detected, and how unwanted change is rejected.

It is not just methods that have developed over time, there has also been significant advances in the quality and availability of earth observation data. The literature of multi-spectral land cover change detection is covered in roughly chronological order and grouped into the three main approaches. The chapter concludes with a discussion of where contributions can be made to advance the literature toward the goal of full automation and high accuracy.

3.2 Approaches to Land Cover Change Detection

3.2.1 Image Pair Analysis

The earliest applications of multi-spectral, satellite based change detection predate the MODIS and AVHRR platforms and made use of data from the early Landsat missions. Due to the low revisit time and the limited availability of images the approach typically taken was to consider only two images and identify signs of change occurring between the two image dates. This approach will be referred to as *image pair analysis*. To minimise the influence of seasonal factors it was important to compare two images taken at roughly the same time of year.

The review paper of Singh [26] provides an excellent starting point for a review of land cover change detection methods published prior to the year 1990. It covers the early approaches to detecting change from Landsat image pairs which includes image differencing [28], image ratios [28]–[30], principal component analysis [31], change vector analysis [32] and post classification change detection [30]. All of these methods, with the exception of the post classification methods, involve transformation of data from pairs of co-registered pixels to produce a suitable metric that can be thresholded to produce a change/no-change decision. Post classification methods will not be considered in this review as they more closely relate to the field of land cover classification.

Subsequent developments in image pair analysis progressed in tandem with developments in satellite sensors. Fung et al. [33] conduct a study to assess the performance of the Thematic Mapper (TM) instrument on board the Landsat 4 platform. This instrument was a major advancement for earth observation adding an additional set of bands designed specifically for vegetation mapping as well as operating at a higher spatial resolution. For the case study of general change detection across a large region of Ontario, Canada, the difference images for the TM bands 3 and 4 (red and near-infrared) were found to be most effective overall. Different types of change could be detected by thresholding different band combinations.

Bruzzone et al. make several major advancements in the field [34], [35]. A method is proposed to reject detection errors caused by image mis-registration or changes in illumination. They make the observation that while land cover changes (forest fires and snow in the case of this study) result in observable changes in only a few bands, sensor

mis-registration errors cause changes that appear uniformly across all bands. Taking advantage of this in a statistical framework the method is able to better discriminate real change from detection errors and as a result achieve significantly lower false alarm rates [34].

All of the previously mentioned image pair analysis methods produce a change metric per pixel which is thresholded to produce a change/no-change map. The threshold is either manually chosen or estimated under the assumption that change is an infrequent event and therefore occurs in the tails of the distribution over all change metric values. The work of [35] presents the first in depth study of the change metric image, the image produced by independently deriving a change metric for each pixel, and proposes two methods which can be applied to these images to calculate a threshold in an unsupervised way. Both methods estimate the densities for change and no-change pixels as well as the prior probability of change and no-change by applying Expectation maximisation (EM) to the change metric image. In the first method these are used directly to estimate a threshold that minimises the probability of error under the assumption that all pixels are spatially independent. The second method takes advantage of the spatial correlations present in the image using a Markov Random Field (MRF) approach [36]. The MRF model allows encoding the prior belief that neighbouring pixels likely share the same change/no-change label. It is then possible to calculate a labeling for the entire image that minimises the expected error under this assumption. From extensive experimentation on real and synthetic data it was found that the first method was able to automatically select a threshold that was very close to optimal for the given data set. This represents a significant step toward the goals of total automation and can be applied to any method which produces a change metric as an output. Incorporating the spatial information was shown to improve detection performance further at the expense of computational complexity and additional parameters.

Image pair analysis may be a good approach where limited data is available but it has major limitations in identifying the date at which the change occurred. It is only possible to say *if* a change occurred between the two image acquisition dates which may be years apart. For certain applications, such as thematic map updating, this does not present a problem. For other applications of change detection, such as land change surveys or invasive species monitoring, the time-of-change is an important output. An increase in

availability of multi-spectral data enabled the next iteration of change detection methods.

3.2.2 Aggregation

The next general approach to change detection will be referred to as *aggregation*. It is characterised by having more than two images but also not taking advantage of the full temporal resolution of the data. Multiple acquisitions are aggregated into their respective years. This greatly simplifies the problem by effectively eliminating the seasonal variations. The trade-off is limited resolution at which the time-of-change can be identified yearly resolution.

The important paper by Lunetta et al. [37] showed the capability of moderate resolution (250m) data from the MODIS sensor for mapping land cover changes at a higher spatial resolution than previously possible with the AVHRR sensor but with availability of many observations per year.

The proposed method operates by calculating the cumulative NDVI (see (2.10) for definition) for each pixel per year. The individual NDVI samples across a year were accumulated and the cumulative NDVI was compared between consecutive years. Change was declared if a significant difference detected. The distribution of differences between non-change years was found to be approximately normal. This was used to selected a threshold such that the probability of change was consistent with known prior probabilities for the particular region and change type.

The strength of this method is in its robustness to natural phenological changes that occur throughout the year. This is something the authors understood well from previous studies using the Landsat TM sensor [38]. By aggregating over a year any variations that occur on an annual basis can be ignored. Combining a years worth of data also has the effect of increasing the signal-to-noise ratio; however, it does have several major drawbacks. Most obviously the temporal resolution at which the change can be located is limited to knowing only the year. For many applications this is entirely adequate. It also does not account for any form of natural variation that occurs on an inter-annual time scale. It is not known if these types of variations are significant in the region of study. Despite these drawbacks the method is shown to perform well in the investigated problem using data from the years 2001 to 2005.

Boriah et al. [39] make an interesting observation regarding the failings of applying change detection methods popular in control theory and signal processing to earth observation data. Like the previous work they identify the issue of seasonality, but also note these kinds of methods fail to take advantage of spatial autocorrelation that exists in earth observation data. To paraphrase, under no change conditions observations that are close in space and time should be similar in appearance and this information could be used to assist in change detection. While they make this observation it is not included in the proposed method and left as a possible avenue for future study.

The proposed method in [39] is a direct successor to the work of [37]. Similarly, MODIS vegetation index data is aggregated on a yearly basis to eliminate the effects of seasonality. The measure of similarity between different years is a distance metric that considers each year as a point in n-dimensional space, where n is the number of samples per year. A recursive merging approach is taken where the most two closest points in space are merged and replaced by a point equidistant between them. This is repeated until only a single signal remains. The final change metric for a pixel is the ratio of the maximum and minimum distance found through this process. This process aims to be more robust to inter-annual variation. The method was applied to the same data set as [37] and through a qualitative assessment found to be more sensitive to certain types of change.

3.2.3 Time Series Analysis

The next major approach to change detection is characterised by the inclusion of many more data points for each location. This became possible with the accumulation of long term data sets from the Landsat, AVHRR and MODIS missions. With many more data points methods from signal processing, regression and time series analysis can be applied successfully and the identification of the time-of-change limited only by the temporal resolution of the data. This section is split into methods that consider only a single image per year, those that consider multiple per year and finally methods that operate online.

3.2.3.1 Yearly Sampling

Kennedy et al. [40] make significant contributions to the field with a particular focus on moving toward total automation in monitoring of changes and disturbances to forests.

They make use of imagery from the Landsat TM sensor and produce a time series with a single cloud free acquisition per year between the years of 1984 to 2004. Only the short-wave infrared band (band 5) was considered. Four different curve models, so called temporal trajectories, are fitted to these time series with each representing a different type of change or disturbance. A step model represents a rapid and permanent change and a step followed by an exponential decay models a change followed by re-vegetation. There are also two models that fit the case where the disturbance occurred prior to the start of the time series: a re-vegetation decay curve and a re-vegetation to stable state. There is no need to model seasonal oscillations as the image is taken from roughly the same date each year. A hypothesis testing approach was used to selected the best fitting model and compare it to the null hypothesis of no-change.

The method can operate in a fully unsupervised manner without requiring the specification of any parameters aside from the significance level to reject the null hypothesis (set to 0.05). From a comprehensive evaluation on forests in Oregon, USA it was found to produce 90% agreement with the results of manual interpretation. The method also produces additional information regarding the year, type and magnitude of change which can be used for subsequent analysis.

The success of this approach lead to the development of Landsat based detection of Trends in Disturbance and Recovery system (LandTrendr) [41], [42].

A supervised time series approach is presented in [43]. They propose training a probabilistic classifier (Random Forest) at each time sample to estimate time series of class membership probabilities. These are monitored for change using the MODTrendr Algorithm [41], a modification of LandTrentr algorithm for MODIS data. As a supervised method this requires a large number of labeled no-change time series to fit the probabilistic classifiers. It also results in a large number of time series to monitor for change if there are many classes being considered.

3.2.3.2 Sub-year Sampling

Considering time series of multiple images per year is the logical progression of change detection as satellite sensor technology advances. It allows for identification of the time-of-change at far greater temporal resolution and methods that operate online, updating their estimates as new data becomes available. With this comes a significant increase in

complexity due to the seasonal oscillations that dominate most land cover time series, especially vegetation. The seasonal oscillations coupled with longer term oscillations and trends have the effect of making the time series highly non-stationary. This is an important theme in land cover time series analysis that is considered in depth in later chapters 4 and 5. Verbesselt et al. articulate this very clearly in their work:

> "Estimating change from remotely sensed data series however is not straightforward, since time series contain a combination of seasonal, gradual and abrupt ecosystem changes occurring in parallel, in addition to noise that originates from the sensing environment (e.g., view angle), remnant geometric errors, atmospheric scatter and cloud effects. The ability of any system to detect ecosystem disturbances depends on its capacity to differentiate the normal phenological cycle from abnormal change [44]."

The majority of methods in this section deal with the problem of transforming the non-stationary time series in to a form that is stationary under no-change conditions. This results in a much simplified problem for discriminating between natural land cover dynamics and change.

A common pattern is to apply a feature extraction method within a temporal sliding window. A suitable feature extraction method should produce similar values under no-change conditions irrespective of the location of the window.

Salmon et al. [2] apply a Fourier transform to the sub-series contained within each window to yield a feature vector comprised of the magnitudes of the Fourier components. It is shown that certain components of the Fourier transform are largely invariant to window position in a no-change time series, but sensitive to changes in land cover type. In this application change detection is performed by clustering the feature vectors at each time step assuming that cluster statistics remain constant through time. Change is detected as a persistent transition between clusters. The method is applied to detecting settlement expansion in Limpopo Province, South Africa. This particular problem and data set appears multiple times within the literature and also within the work of this thesis. The work of [45] similarly applies the Fourier transform for feature extraction but instead applies a supervised classification approach (multi-layer perceptron) to estimate class membership directly.

Another dominant approach in the literature for dealing with non-stationary time series involves fitting a parametric model to a time series and assessing the model parameters to identify change. These types of methods impose strong assumptions that the data follows a particular parametric model. For change detection in land cover time series an annual periodic model is almost always used possibly with additional harmonics or trend. Ronald et al. [46] demonstrated an early approach to model parameter assessment in land cover time series. For each year a harmonic model was fitted using least squares. The model takes the form

$$x(t) = \mu_0 + \sum_{n=1}^{n=N} \alpha_n \sin\left(\frac{2\pi n t}{T} + \phi_n\right) \tag{3.1}$$

«««< HEAD where x is the observed pixel value, μ_0 is the bias, α_n and ϕ_n are the amplitude and phase of the n^{th} component and T is the sample period. $N = 2$ components were used. This is almost equivalent to the Fourier transform method of [2] however fitting the model using least squares is more robust to irregular samples and missing data [46]. In this study the trend of the bias and amplitude parameters across several years were used as a visualisation tool for mapping. ======= where $x(t)$ is the predicted signal, μ is the bias, α_n and ϕ_n are the amplitude and phase of the n^{th} component and T is the sample period. In [46] only the seasonal component plus one additional harmonic ($N = 2$) is considered. This is almost equivalent to the Fourier transform method of [2] however fitting the model using least squares is more robust to irregular samples and missing data [46]. In this study the trend of the bias and amplitude parameters across several years were used as a visualisation tool for mapping. »»»> reviewer-1-fixes

Verbessalt et al. [47], [48] apply parametric models to detect change in vegetated time series. Rather than using a windowing approach, change points are identified by allowing the model to *break* in to two pieces which are fitted independently. The break points are selected to minimise the sum of squared errors. This method can be applied recursively to identify multiple break points in a time series. This approach, termed Breaks For Additive Seasonal And Trend (BFAST), allows for the identification of the time-of-change as well as some characteristics of the change point such as its magnitude. In [47] the seasonality is accounted for by including a dummy variable that encodes the time of year. The coefficients of the dummy variable are constrained to sum to one over

a yearly cycle. This model is flexible and can fit non-sinusoidal oscillations. In [48] a harmonic model of (3.1) with $N = 3$ is used to encode the seasonal cycles.

Time series were first classified as change/no-change using the Ordinary Least Squares Moving Sum (OLS-MOSUM) test [49] which provides a method for detecting structural changes from a model based on residuals and parameter estimates. This was applied independently to the trend and seasonal models of the time series. If a time series was classified as containing at least one change the position of the breakpoints were estimated. The result is an estimate of the number of changes, along with a piecewise model providing useful information for the changes in each time series. Both method were applied to real and simulated MODIS NDVI time series to detect forest disturbances in Australia. The harmonic model was found to be more robust and resulted in better change point estimates. Through detailed analysis it was shown that the accuracy of the method in detecting change locations is a function, not only of change magnitude to noise ratio, but also to seasonal amplitude with a larger seasonal amplitude yielding better detection accuracy. An estimate was provided of the signal-to-noise ratios (SNR) and amplitude to noise ratios under which the proposed method can be expected to work well.

A Kalman filter allows for the estimation of parameters in a linear system that evolve over time without the use of a temporal window. There are also non-linear variants, such as the Extended Kalman Filter (EKF) [50], that allow this to be applied to non-linear models such as harmonic model of (3.1). This is the method proposed by Kleynhans et al. in [6]. The harmonic model of (3.1) with $N = 1$, termed a triply modulated cosine, is fit dynamically using an EKF. The resulting time series of parameters is used to detect change by comparing the time series with those of neighbouring pixels and declaring a change when the deviation of a time series from its neighbours exceeds a threshold. Interestingly, although the Kalman filter method operates online the final change decision process is offline as the samples from the entire series are used when calculating the change metric. The change metric is designed under the assumption that the time series of parameter values should be stable under no-change condition but experience a shift during land cover conversion. This was observed by analysis of the sum of the derivative of the parameter signals. This should be close to zero under no-change but accumulate a bias if a change occurs.

The change metric in [6] also makes use of spatial information in the form of neighbouring pixels when calculating the change metric. The final metric is defined as the total distance between the sum of the derivative signal for the center pixel and its direct neighbours. The inclusion of spatial information aims to alleviate errors due to inter-annual variations. It is assumed that all the pixels in a close neighbourhood would have similar land cover and experience similar climate, thus vary similarly between years. It also assumes that change moves gradually across the landscape so that as the change moves across the neighbourhood of interest there will develop a disparity between the parameter series of adjacent pixels. These assumptions appear to hold well for the problem assessed in the paper, detecting settlement expansion in South Africa using MODIS time series, as the natural land cover is spatially homogeneous at the sensor resolution and change typically progresses outwardly from existing settlements. For other problems where adjacent pixels are dissimilar and change occurs rapidly over large regions, such as fire, the proposed method of including spatial information could degrade performance.

The extended Kalman filter method influenced several later works. Salmon et al. [51] define a new change metric based on the internal covariance matrix rather than parameter estimates. The rationale behind this is that the covariance matrix of the EKF drives the changes in the parameter vector, thus monitoring it directly may improve detection. Also in this case no spatial information is included, the change detection metric is defined as the maximum absolute deviation of a given covariance matrix parameter from its expected value under no-change conditions. They also consider a multivariate change metric that makes use of multiple spectral bands and multiple model parameters simultaneously. In this case the change threshold is estimated using a supervised learning approach. Both approaches were found to be superior to assessing the parameters of the EKF only and the multivariate case was found to be most effective for change detection overall.

Something not discussed so far is the large number of parameters that must be selected for an EKF (initial parameter vector, state transition matrix, observation matrix) and the sensitivity of the method to these parameters. Parameter selection is challenging for operators and requires accurately annotated change/no-change data sets sets to validate against which detracts from the goals of total automation. The work of [52] aims to address this by applying an optimisation procedure to select the parameters for the EKF

in an unsupervised way.

Subsequent papers by Salmon et al. employ the EKF as a feature extractor [1], [3]. The features are assessed for land cover change using a novel method based on non-linear systems theory. Notably all of the published methods making use of the EKF for land cover change detection use offline detection techniques despite its potential as an online method.

An alternative approach to change detection in non-stationary time series is to calculate some metric that quantifies the degree of non-stationarity. Change can be detected under the hypothesis that time series experiencing change should have a higher degree of non-stationarity than no-change time series. Kleynhans et al. [5], [7] accomplish this by calculating the autocorrelation under the wide-sense stationary assumption, which assumes that the mean and variance at a certain lag remain constant though time. A greater discrepancy between the overall signal mean/variance and the instantaneous measurements causes a greater drift in the autocorrelation function from its stationarity form. This drift is then thresholded to yield a change/no-change decision.

The method is improved upon through incorporation of spatial information, automated selection of the lag parameter and inclusion of multiple spectral bands [53]. The lag parameter was eliminated by using the sum of all lags when calculating the change metric rather than a single lag. The inclusion of spatial information and multiple bands was done in a similar way to [6]. The improved change metric was calculated as the euclidean distance between the center pixel and the mean of the pixels within a fixed size spatial window. This distance was calculated in the 7-dimensional space comprised of all MODIS bands. Results directly compared with [7] show that despite a reduction in the parameters this method is able to decrease the false positive rate for detecting settlement expansion from 17% to 1% without sacrificing detection accuracy.

3.2.3.3 Online Methods

Online change detection methods are characterised by their ability to update their change/no-change decision as each new data point becomes available. This is done with the intention of detecting change as quickly as possible. These methods require a three way trade-off between detection accuracy, false alarm rate and detection delay.

Any offline change detection method can be converted to an online method by applying it in a sliding window that advances as data becomes available [54] and as such any of the methods above could be converted to operate online.

This is the approach taken by Anees et al. toward the problem of detecting bark beetle infestations in British Columbia, Canada. In [55] the parametric model approach of [46] is applied in a sliding window and the resulting parameter time series are monitored online for change. This is done by assuming that the differences between consecutive mean parameter estimates is zero-mean Gaussian under no-change conditions. The variance is estimated for each signal during a learning phase and used to produce a series of z-scores, a derived statistic which should be zero-mean, unit variance Gaussian under no-change conditions. A change alarm is declared if a z-score exceeds some threshold. This method was evaluated on various MODIS vegetation indices and was found to yield good true-positive/false-alarm performance for an optimally selected threshold, albeit with a mean detection delay of 560 days.

The method was improved upon in [9] by modifying the conditions under which the alarm is declared from a single value exceeding the threshold to a heuristic that requires more than six values exceeding the threshold within the previous ten. This aims to reduce the sensitivity of the method to singular outlying samples, allowing the threshold to be decreased while maintaining a similar false alarm rate. A comprehensive study on real and simulated data was included and true-positive/false-alarm performance was shown to be improved while detection delay remained greater than 500 days when applied to real data and around 400 for simulated data.

In the majority of cases a sliding window is likely not the best approach. Offline detection methods are often insensitive to change occurring at the edges of the signal thus the sliding window must advance substantially beyond the change point before the change can be detected. This is supported by the long detection delays in [9], [55]. For certain applications that require the change be detected as rapidly as possible these delays are unacceptable.

Anees et al. abandon the windowing approach in favor of the window-less Kalman filter method of [6], [51] to estimate parametric model parameters online [56]. Furthermore the time series of estimated mean parameters is monitored for change using a more statistically rigorous approach (CUmilative SUM control charts (CUSUM) [57], [58]) than

the heuristic of [9]. The threshold for the CUSUM is selected in an interesting way that minimises a cost function which takes in to account the three major performance metrics: true alarm rate, false alarm rate and detection delay.

Chandola et al. [59] detail a method designed for rapidly detecting change in periodic, non-stationary time series, not limited to land cover time series. It operates by using Gaussian Process regression to make a probabilistic forecast of the next sample given the historical samples in a time series. Once the next sample is available its z-score under the forecast distribution is calculated. By repeatedly applying this forecasting method a time series of z-scores is produced. Provided the model is working correctly this time series of z-scores should be Gaussian i.i.d. A series of samples that deviate significantly from the forecast will result in a change in distribution of the series of z-scores. This is detected using an Exponentially Weighted Moving Average (EWMA) control chart [58]. Much like the parametric model approaches, some prior knowledge about the structure of the time series under no-change conditions is included. For a Gaussian process this kind of knowledge is encoded in the covariance function. A periodic exponential covariance function was selected which encodes the prior belief that samples which are temporally close are correlated, as are samples that are close to integer multiples of the period.

The kernel function requires the selection of five hyper-parameters which can be considered as statistical properties of the time series. For each series these were estimated during an unsupervised learning phase and then held constant. The method also requires the selection of two parameters to control the sensitivity of the EWMA control chart. The method was applied to detecting change in MODIS NDVI signals of crops in Iowa, USA. It was compared with the offline approach of Boriah et al [39] as well as an alternative approach for online detection in periodic time series using a Seasonal Auto Regressive Moving Average (SARIMA) model [60]. The time signals periodicity and trend was removed and the residual used to compare with other online change detection methods that assume stationarity including CUSUM control charts [57], [58] and Bayesian online change detection [61]. The Gaussian process approach was found to yield the highest detection accuracy, recall and precision. No comparison was made for detection delay. The proposed approach was highly novel however it requires a total of seven parameters, to each of which the performance is very sensitive (see Chapter 6). The computational performance and scalability of the method was improved upon in [62].

Verbesselt et al. [44] propose a modification of their earlier work [47], [48] for online detection. The same seasonal and trend model is used but it is fit on a window of data assumed to have experienced no change: the *history period*. The model is then used to predict the values in a window of values immediately prior to the most recent sample: the *monitoring period*. Under the no-change hypothesis the parameters estimated from the history period should provide good estimates for the monitoring period. If a structural change has occurred this is not the case. The change hypothesis is tested by computing the MOSUM [49] of the residuals in the monitoring period and thresholding them at a certain significance level. This approach differs in its implementation of online detection compared with the other online methods considered [4], [59]. Rather than attempting to detect change with the smallest possible detection delay in each case, the minimum detection delay is fixed in advance as the length of the monitoring period. This could be considered as an intermediate between online and offline methods. In applications where there is no advantage to achieving a detection delay less than some minimum value this could be a useful approach to reducing the false alarm rate.

Grobler et al [4] use a supervised learning approach to convert a time series of MODIS reflectances into time series of class membership probabilities. Class specific densities are estimated for each time step using a collection of labeled training series. The log-likelihood ratio is calculated from observations and monitored to detect a transition from one class to another using a CUSUM control chart [57], [58]. This approach makes very few assumptions about the structure of the data. As densities are estimated independently at each time step seasonal variation is accounted for. In order to obtain more training data for each time step the data from consecutive years were aggregated. As a result inter-annual variations were not accounted for and the time series were assumed to be cyclo-stationary. This was, however, a choice made to obtain more training examples per sample and not a requirement of the method itself.

Reiche et al. contribute several works that combine multiple data sources in order to detect change. In [63] Landsat NDVI data is combined with Synthetic Aperture Radar (SAR) data from the ALOS PALSAR platform for online detection of deforestation. This has the advantages of increasing the effective temporal resolution, alleviating the problem of cloud cover and providing a complementary perspective of the land cover. The data is fused using a probabilistic graphical model approach. A training set of forest and

non-forest observations is used to estimate an independent Gaussian density for each time sample and for each sensor, similar to [4]. At detection time a Bayesian updating procedure is applied to iteratively calculate the probability of deforestation having occurred given the observations to date. A change alarm is triggered when the probability of change given the observations exceeds a given threshold. Fusing multiple sensors in this way was shown to consistently outperform using either sensor alone, both in terms of detection delay and accuracy.

[64] extends the method further to also include data from the Sentinel-1 SAR platform. This is applied to detecting deforestation in tropical areas where pervasive cloud cover prevents consistent multi-spectral observations. This method also includes a spatial normalisation step in order to eliminate the natural variations in the signal. This allows for only a single density to be estimated for forest/non-forest per sensor, rather than one for each time step as in the previous method. Again this method showed improvements in delay and accuracy when combining data sources, in particular the Sentinel-1 SAR data was found to be very effective.

Both of the above methods demonstrate excellent advancements in multi-sensor fusion for change detection. It is clear that this is a promising strategy for dealing with the problems of missing data and cloud cover while also providing additional information from which to derive change decisions. Being supervised methods both require training sets of forest/non-forest time series. It is also not clear how they could be generalised to change between unknown classes, an issue shared with [4].

3.3 Conclusion

Throughout the time period considered in this review there have been major advances both in satellite sensors and change detection methodologies. The most recent works make use of time series methods applied to high temporal resolution data. Every method considered takes a particular approach to estimating and rejecting natural variations and changes, while remaining sensitive to the types of change that are of interest. This is at the heart of the problem of land cover change detection. Perfect knowledge of the natural land cover dynamics for a particular region permits error free change detection. This is

particularly true for online change detection where the natural changes must be identified and rejected in real-time based only on historical data. It is clear that improvements in methods for estimating the natural land cover dynamics are essential for advancing method for land cover change detection.

Another common theme throughout the literature is taking advantage of spatial and spectral correlations to define robust change thresholds. This was shown in several works to improve change detection accuracy while incurring minimal additional costs in terms of complexity and free parameters. It is clear there is an abundance of information available by considering spatial correlations and multiple bands and these should be taken advantage of where possible. There is also the possibility of fusing data from multiple sources to detect change, a recent development made possible by the abundance of satellite sensors. While this poses new challenges for fusing different temporal and spatial resolutions it has been shown to be effective, especially in places where consistent observations from a single sensor are not possible.

Finally there is the problem of automation. In general, as time has progressed the published methods have decreased in their overall levels of automation from unsupervised image pair analysis methods that require either a single threshold or even no parameters, to supervised learning approaches requiring multiple parameters and a training set. This is expected as the complexity of algorithms increase but it is imperative that these parameters be kept to a minimum and unsupervised algorithms designed where possible to allow them to be run by non-experts.

Advancing the field of land cover change detection will require taking advantage of all of the potential sources of information of previous approaches. Temporal and spatial correlations have been shown to contain much of the information needed to discriminate change and no-change time series without requiring supervision. It is also clear from several studies that including more than a single spectral band or even more than one sensor has the potential to improve change detection performance both in the offline and online paradigms. What is needed is a general framework under which all of this information can be rationalised in order to make a decision.

Chapter 4

Modeling Land Cover Dynamics for Change Detection

4.1 Introduction

The previous chapter concluded the need for a general framework under which temporal and spatial information can be combined to infer if a particular time series has experienced land cover change. Such a framework must be capable of modeling the natural phenological variations present in land cover time series so they can be rejected for the purpose of detecting the change types of interest. In this section a model for offline detection of change in fixed length time series is considered.

For this purpose a probabilistic approach of estimating the joint Probability Density Function (PDF) of samples in a time series is considered. Such a density function has dimensionality equal to the length of a time series and encapsulates the probabilistic relationships between each sample and every other sample. Such an approach has not been considered in the literature with methods instead preferring to assume that samples are either temporally independent [4], [43], [63], [64] or follow a particular parametric model [6], [44], [47], [48], [55], [56]. One exception is the work of Chandola et al. [59], [62] which directly models the joint PDF as a continuous Gaussian process.

In this chapter it is argued that, as land cover time series are discrete time processes with a fixed sample rate, a continuous time Gaussian process is not an ideal choice. Instead the discrete time equivalent, a multi-variate Gaussian PDF, is more appropriate. It is also argued that the parameters for the joint PDF can be estimated in an unsupervised manner from a collection of time series within a spatial neighbourhood. Such a model

should capture the natural variations in a region in a probabilistic sense, allowing change to be declared when a particular time series does not follow its expected trajectory. This leads to the first hypothesis:

> *Time series experiencing change will appear as outliers relative to a multivariate joint PDF fit from a collection of land cover time series in a geographical region.*

Meeting this hypothesis satisfies the first two requirements for a general framework, incorporating temporal and spatial information. The limited number of free parameters for such a method also contributes toward the goal of total automation.

Another important aspect to the problem is identifying the time-of-change, known as change point estimation. A method is also presented for using the time series joint PDF for change point estimation. Under the hypothesis that correlations are weaker between samples in a time series before and after a change point, the problem of change point estimation is formulated as a multiple hypothesis test. Each hypothesis is that the change occurred at a particular sample and the correlations before and after that point are zero. This leads a second, similar hypothesis for this chapter:

> *Enforcing in the model that the correlations between samples before and after a change point are zero can improve the fit which can be used to estimate the time of change.*

This chapter begins with an exploration of the idea of correlation between samples in time series. This is extended to develop an unsupervised method for change detection. This follows on to a discussion of different approaches to using a joint PDF to estimate the time of change. Several methods are considered for testing. The chapter concludes with the design of series of experiments test the above hypotheses. One important consideration in these experiments is the selection of a set of appropriate methods from the literature with which to compare, as well as suitably informative metrics with which to compare them.

The content of this chapter has been compiled into a publication that appears in IEEE Geoscience and Remote Sensing Letters. See Appendix D or [11].

4.2 Background

When considering offline change detection over a given spatial region it can be expected that there will be a large collection of time series available, all of which are of the same fixed length, taken from the same sensor and sampled at approximately the same time. Therefore given two spatially close no-change time series from identical patches of land it can be expected that the time series would be identical except for noise induced by the observation system as discussed in Chapter 2. Given enough identical patches it would therefore be possible to estimate independently, at each point in time, the expected reflectance for that type of land cover and the system noise as the mean and variance of the measurements respectively. This is the approach taken by some of the supervised methods in the literature [4], [43], [63], [64] under the assumption that patches of the same land cover class are almost identical. A separate density is estimated for each class and any within class variance is factored in with the noise.

Let $\mathcal{X} = \{\mathbf{x}^{(i)}\}_{i=1}^{i=D}$ be a set of pixel time series for a single band captured over a geographical region within some fixed time span of length D. Each time series is a vector $\mathbf{x} = [x_1, \ldots, x_N]^T$. Each series can be considered an observation of a random variable $\mathbf{X} = [X_1, \ldots, X_N]^T$. Given a set of unlabeled time series a naive approach to estimating the natural land cover dynamics of the region might involve calculating the mean signal from the set. A simple probabilistic approach might assume that each observation is generated from this mean signal plus some Gaussian noise that encapsulates the differences in vegetation as well as any noise from the observation system. This model can be expressed as

$$X_k \sim \mathcal{N}(\mu_k, \sigma_k^2) \tag{4.1}$$

with the mean, μ_k, and variance, σ_k^2, estimated independently for each time sample. The probability of observing a time series under this model is given by

$$\Pr(\mathbf{X} = \mathbf{x}) = \prod_{k=1}^{N} \Pr(X_k = x_k). \tag{4.2}$$

Fig. 4.1a shows the above model fit from a collection of land cover time series of native vegetation in rural South Africa [65]. The mean signal gives an estimate of the natural land cover dynamics of the particular vegetation class in the region while the model

uncertainty is shown as the 95% confidence interval.

Such an approach makes strong oversimplifications about the nature of land cover. Firstly, land cover classes themselves are notoriously difficult to define [66]. In reality no two patches of land cover are identical. They are comprised of mixtures in varying fractions of different types of vegetation, water, soil, rock and possibly anthropogenic artifacts. The reflectance measurements obtained by the satellite system for a given patch are therefore a non-linear function of these factors as well as the terrain geometry. Unsurprisingly the within-class variance can still be significant.

Rather than treating each sample in a time series as being generated by some univariate class conditional probabilistic process an alternative approach is to consider each time series as a single observation from a multi-variate process. Under this model the probability of observing a signal is

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{4.3}$$

This model is the discrete equivalent to a continuous time Gaussian process. There are several advantages to this approach. Firstly, much like a Gaussian process, this model is able to consider the correlation between samples in the same signal. Unlike a Gaussian process the correlations need not be specified by a parametric covariance function but can be estimated directly given a number of time series with the same covariance structure. A number of time series of similar land cover from a particular geographical region should indeed share similar correlation structure and that this should encapsulate the phenological and climatic changes of that region.

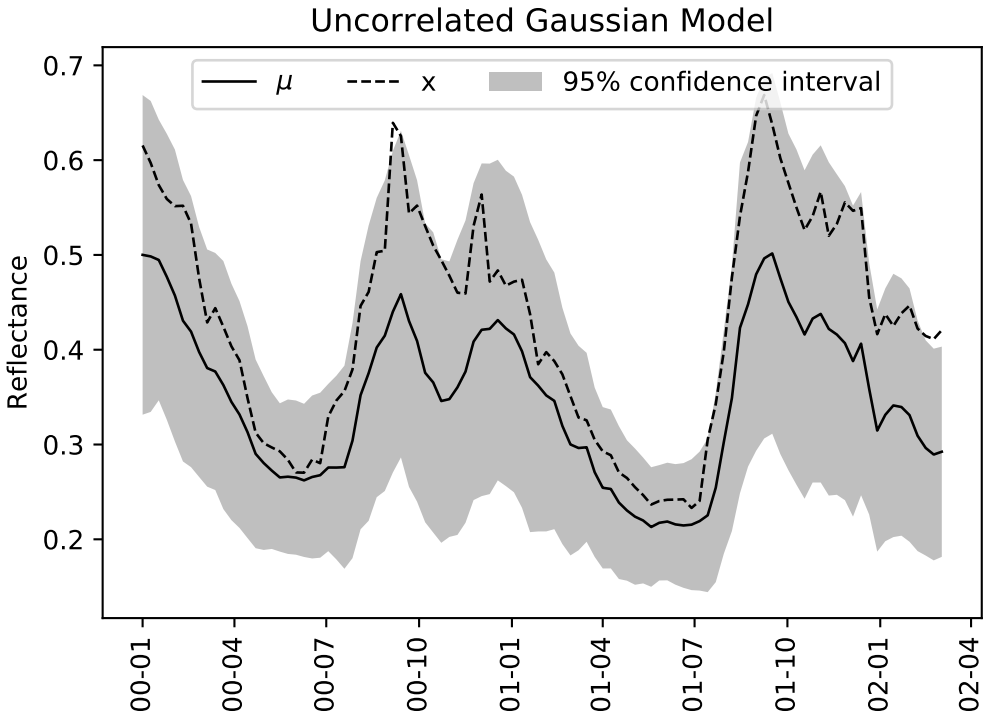
Fig. 4.2 gives an example of two MODIS pixel time series from the Limpopo Province, South Africa which will be used to rationalise the above statement. These pixels are spatially close but contain different types of vegetation with one being dominated by grasslands and the other by small shrubs. These two land cover types are representative of the majority of the vegetation in this area. From the time series it is possible to make three several interesting observations. The time series for the shrub pixel has consistently smaller NDVI throughout which could be represented by a downward shift on the NDVI axis. The two time series share the same seasonality. Both exhibit a rapid increase in NDVI between October and March which corresponds to the growing season for this

region [65]. The time series share the same inter-annual characteristics. Both exhibit higher NDVI during the summer of 2004 than the summer of 2003 for example. Southern Africa was known to have experienced a drought between 2001 and 2003 [67].

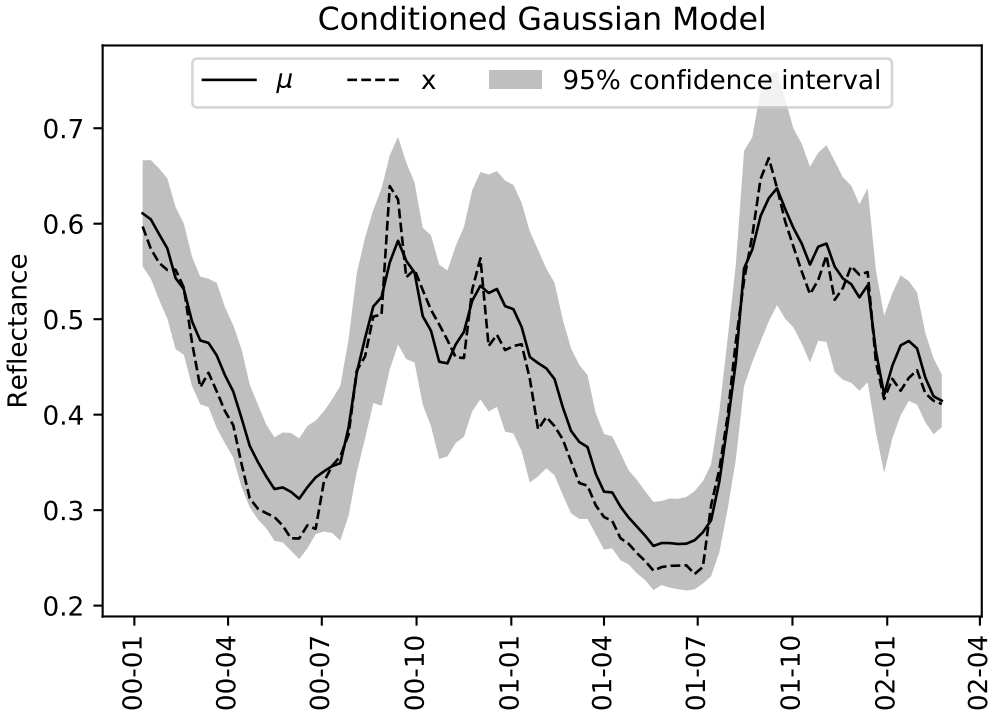
Given enough examples of pixels sharing the characteristics discussed it is possible for a joint PDF to capture these relationships in a probabilistic sense in its mean vector and covariance matrix. Fig. 4.3 shows the two signals from Fig. 4.2 along with the mean vector and correlation matrix estimated from 2348 time series from the same region. The correlation matrix is plotted such that it aligns with the time series plots. From this figure it is possible to make the following observations about the estimated model.

- The mean vector captures the general shape of the seasonality and the differences between the seasons (inter-annual variations) but is off by a scaling factor.
- The uncertainty (shown by the standard deviation vector, σ, extracted from the diagonal of the covariance matrix) is much higher at certain times of year than others
- All correlations are positive implying that time series tend to shift up or down from the mean across the entire time range
- The blocks of higher correlation along the diagonal corresponds to the lowest NDVI sections of the seasonal cycle. These are also correlated with the same seasons in other years as shown by the off-diagonal blocks. These periodic changes in correlation suggest that the low season from previous years gives a better indication of what to expect for the current year than observations from the previous season.
- The peak of the growing season has low correlation with the rest of the signal suggesting it is not a good time to be detecting change.
- The matrix does not show strong banding along the diagonal. This suggests that all of the samples in signal are strongly correlated not just those temporally close.

All of this information encapsulated by the Gaussian model can be used to make inferences about a time series given a subset of observations. For example conditioning on the first and last observations of the signal can be used to improve the estimate of the mean and reduce uncertainty in the remainder. This procedure was applied to the same



(A) Marginal PDF of a Gaussian model that assumes uncorrelated samples



(B) Marginal PDF of a multi-variate Gaussian model conditioned on the first and last observations

FIGURE 4.1: Example for how modeling correlation and conditioning can be used to improve mean estimates and reduce uncertainty. Time series, x, is a real MODIS time series of band 1 reflectance captured for vegetated land cover in Limpopo Province, South Africa. μ is the model mean signal in each case and the 95% confidence interval is derived from the model standard deviation signal. See Chapter 6 for further details regarding the data set

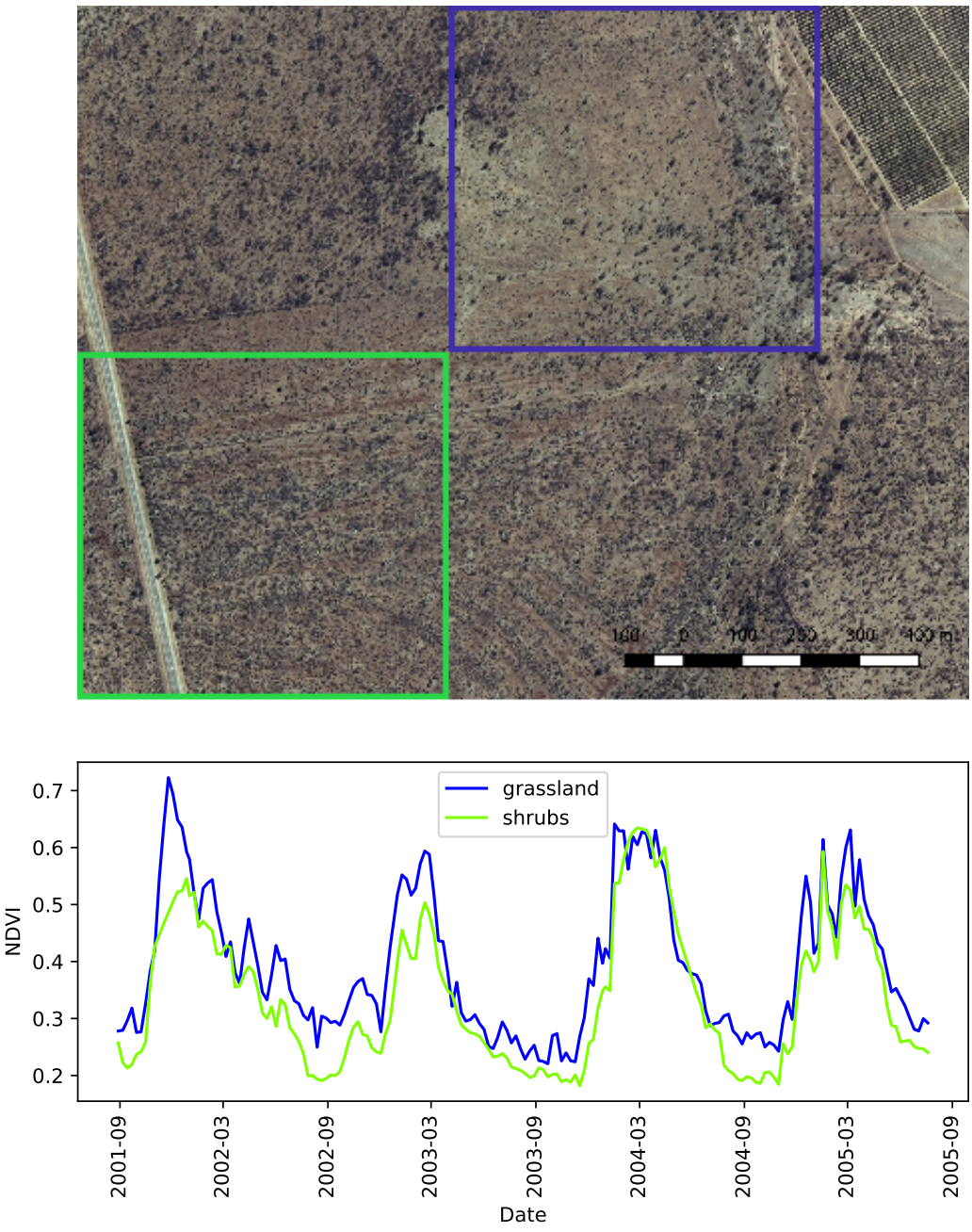


FIGURE 4.2: Two adjacent MODIS pixels from the Limpopo Province, South Africa and their accompanying NDVI time series spanning 4 years.

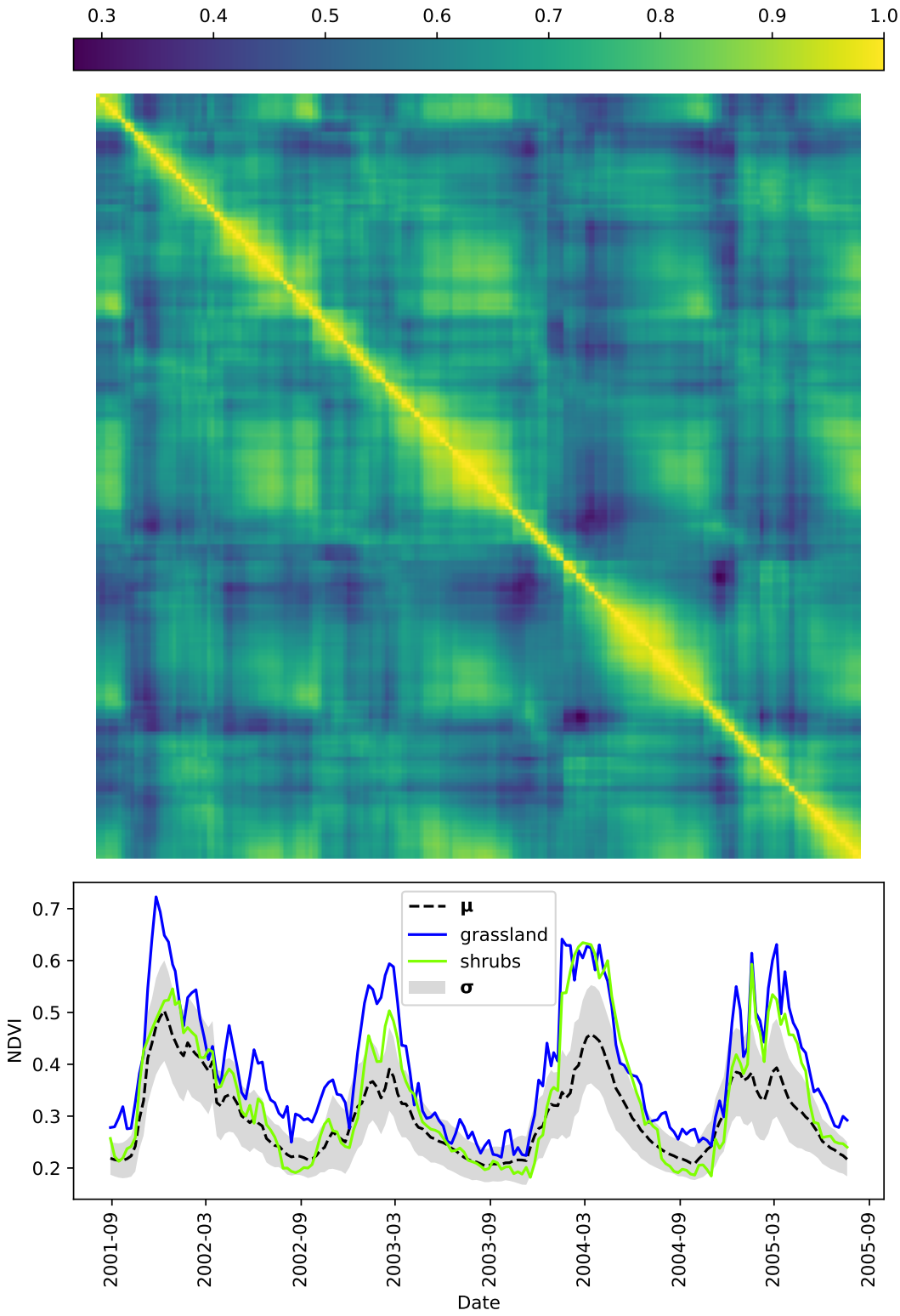


FIGURE 4.3: The correlation matrix (normalised covariance matrix) estimated from 2348 time series from Limpopo Province, South Africa along with the plot of the mean vector and two time series from Fig. 4.2.

time series as in Fig. 4.1a using the Gaussian model estimated in Fig. 4.2. Comparing Fig. 4.1a and Fig. 4.1b it can be seen that these observations significantly effect the estimate of the mean and variance, bringing the mean closer to the observed signal and reducing the uncertainty. It is also worth noting that even though observations are only provided at the beginning and end of the signal the uncertainty is reduced throughout. A Gaussian joint model has the additional advantage that both conditioning and marginalisation of the probability density can be done in closed form.

4.3 Change Detection

Once a suitable probabilistic model of the natural land cover dynamics has been obtained there still remains the problem of how to obtain a change/no-change decision for each time series. This requires the derivation of a suitable metric which can be thresholded to yield a decision. Such a metric should encode the following:

> *Given a candidate time series and a collection of similar time series from the same geographical region, what is the likelihood that that series was generated by land cover that did not experience change?*

The log-likelihood for a sample under a multi-variate Gaussian PDF can be calculated as

$$\ln L = -\frac{1}{2}\left(\ln(|\mathbf{\Sigma}|) - (\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) + N\log(2\pi)\right). \tag{4.4}$$

where N is the dimensionality.

Extracting only the terms that are a function of $\mathbf{x}$ and taking the square root results in the *Mahalanobis distance* which is defined as

$$D_M(\mathbf{x}) = \sqrt{(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})} \tag{4.5}$$

for a candidate time series $\mathbf{x}$. This distance metric is equivalent to the Euclidean distance between an observation vector and the mean vector in the space resulting from a whitening transform of the data.

The Mahalanobis distance can be used to discriminate between change and no-change time series by selected some threshold, δ, such that a change is declared when

$$D_M(\mathbf{x}) > \delta. \tag{4.6}$$

This approach to change detection will be referred to as *Mahalanobis distance change detection*. Such a method is unrepresented in the current literature for land cover change detection but similar methods are commonly used for outlier detection problems [68]. Offline change detection with an estimated joint PDF can be considered equivalent to the problem of detecting outliers in a multi-dimensional data set. Any time series with points that do not follow the expected trajectory will appear as outliers with respect to the joint PDF.

4.4 Change Point Estimation

There still remains the problem of how to identify the point where the change occurred. In the previous section the finite length time series are considered as N-dimensional feature vectors. The feature vectors are used to define points in N-dimensional space and this is used to determine outliers. Such a method is invariant to re-ordering of features/time-samples provided the same re-ordering is applied to all vectors in the training and test set. This means that the *time* aspect is effectively not considered and as a result it is impossible to identify the time-of-change. In this section a method is proposed for estimating the time-of-change given a candidate time series and a joint PDF.

Consider a time series with a single land cover conversion at index k^*. It is expected that values of the signal prior to k^* should be less correlated with values after k^* and visa versa as they are generated by different types of land cover with different spectral properties and phenology. It is proposed that a better fitting model can be obtained by assuming the values before and after k^* to be totally independent, allowing the joint density to be factorised as

$$\Pr_{k^*}(\mathbf{x}) = \Pr(x_1, \ldots, x_{k^*}) \Pr(x_{k^*+1}, \ldots, x_N). \tag{4.7}$$

Without refitting the model this assumption can be imposed by zeroing the off-diagonal

TABLE 4.1: Comparison of average log-likelihoods on synthetic change data

Method	Average Log-Likelihood
Direct Fit	1562
No-change Fit	1314
Modified Covariance	1457

elements of the covariance matrix that relate values before and after the change point. In block matrix notation this can be expressed as

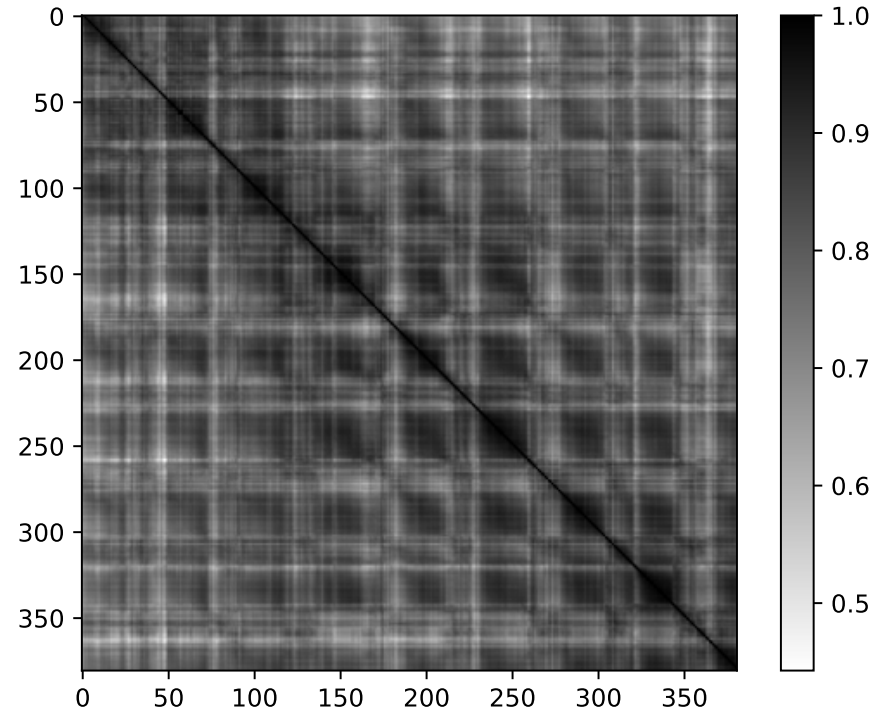
$$\mathbf{\Sigma}(k) = \begin{bmatrix} \mathbf{\Sigma}_{(1,\ldots,k^*)(1,\ldots,k^*)} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{(k^*+1,\ldots,N)(k^*+1,\ldots,N)} \end{bmatrix}. \tag{4.8}$$

This operation will be referred to as *partitioning* the joint density at k.

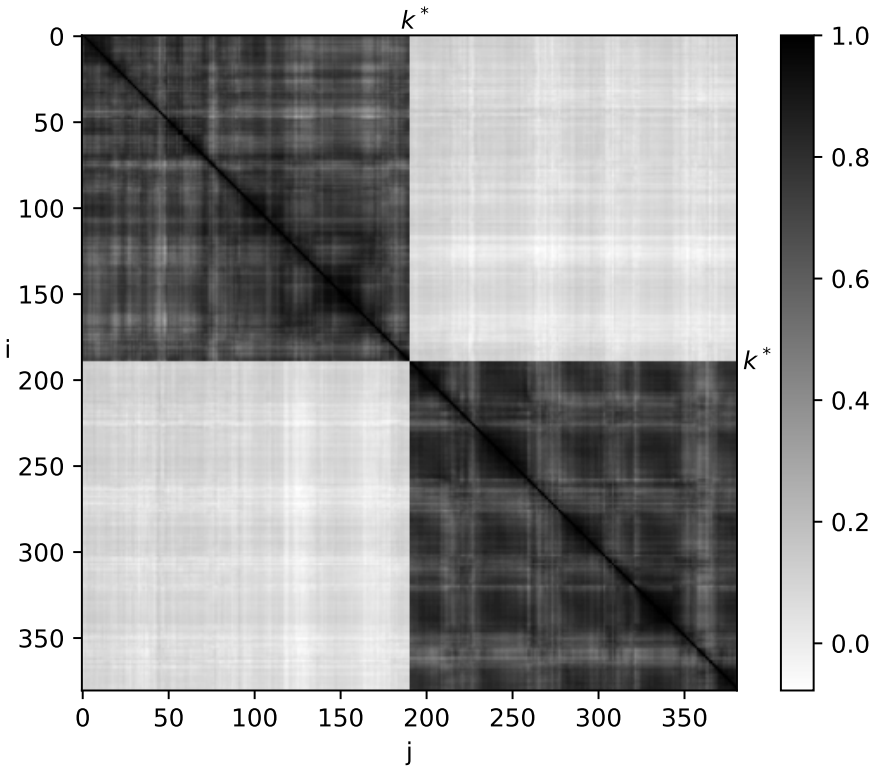
In order to validate that the goodness-of-fit is indeed improved by this modification when change points are present, we compared it against the original model and also a Gaussian model fit directly from a set of change time series. A synthetic data set for this purpose was created by randomly sampling with replacement 10000 vegetation and rural settlement time series from the original data set and concatenating them with an identical and known change point at $k^* = 190$. A comprehensive description of the methods used to produce synthetic change data can be found in Chapter 6. Fig. 4.4a shows the correlation matrix (normalised covariance matrix) estimated from the full data set with majority no-change and Fig. 4.4b shows the same process applied to a synthetic change data set. It can be seen that the correlation between samples before and after k^* in the synthetic change signals are close to zero.

To quantify the goodness-of-fit the average log-likelihood was calculated for the synthetic data set. This gives the expected log probability of observing a time series from the data set under the given model. A higher value indicates a better fitting model. Table 4.1 shows that by enforcing independence post-fit the likelihood approaches that of the model fit directly to the change data.

When the time of change is unknown consider the hypothesis that the signal change point occurred at time index k, denoted as $\mathbf{H}_k$. Under this hypothesis it is expected that the likelihood can be improved by replacing the covariance of the model with that of (4.8).



(A) Correlation Matrix for No-Change Series



(B) Correlation Matrix for Synthetic Change Series with change point $k^* = 190$

FIGURE 4.4: Correlation matrix (normalised covariance matrix) heat maps estimated for MODIS band 1 from signals in the region of interest in Limpopo, South Africa.

Identifying the change point is then a matter of evaluating all possible hypothesis and selecting the one that maximises the likelihood or equivalently minimises the Mahalanobis distance. Under this model the best estimate of the change point is

$$\hat{k} = \underset{1 \leq k \leq N}{\arg\min} \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}(k)^{-1} (\mathbf{x} - \boldsymbol{\mu})}. \tag{4.9}$$

Fig. 4.5 presents an example of this method applied to a hybrid synthetic change time series produced using the method described in 6.2.3. The blended time series were randomly selected from collections of vegetation and settlement type land cover and the combined covariance matrix is the same as shown in 4.4a and was estimated from the combined collection using both the empirical and robust methods. It can be seen that, for a synthetic time series containing a known change point, the model is an increasingly better fit as the proposed change point approaches the actual change point. For a no-change signal the distance does not change significantly.

4.5 Considerations for estimating the joint density

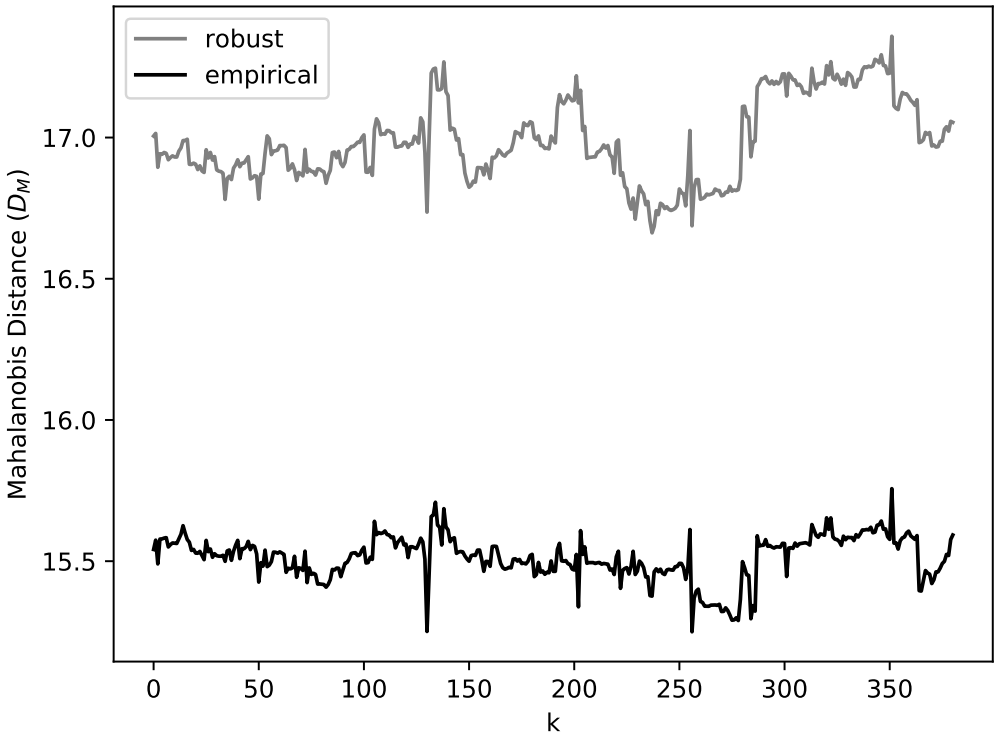
4.5.1 Robust Covariance Estimation

It is well known that the empirical method of estimating a covariance matrix from data,

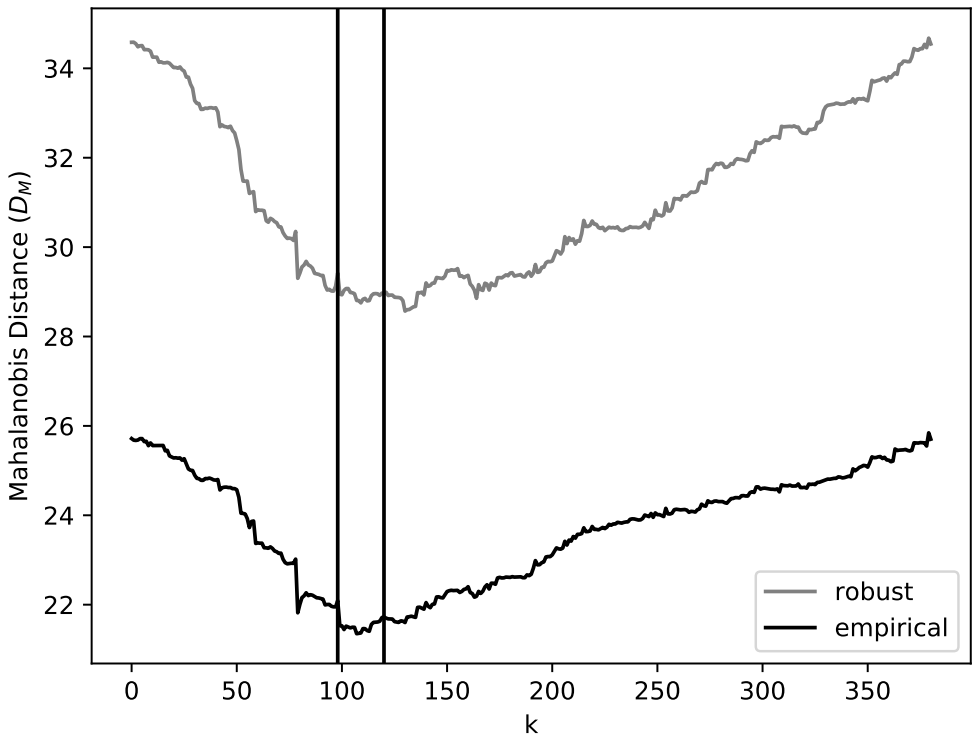
$$\boldsymbol{\Sigma} = \frac{1}{D-1} \sum_{i=1}^{D} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T, \tag{4.10}$$

is notoriously sensitive to the presence of outliers in the data set [69]. Even a single outlier, caused by an error in the sensor system or a particularly unusual region of land cover, could inflate the covariance estimates resulting in reduced sensitivity to detecting changes. As many different applications require the robust estimation of a covariance matrix in the presence of outliers, many different methods have been proposed.

One solution proposed by Rousseeuw et al [69] to estimate the covariance using only h out of the total D examples. h examples are selected as the subset of samples that yields the covariance matrix with the smallest determinant, equivalent to the PDF with the smallest entropy. This is known as Minimum Covariance Determinate (MCD) estimation. While the subset size, h, could be treated as a hyper-parameter it is instead



(A) No-Change Signal



(B) Synthetic Change Signal

FIGURE 4.5: Mahalanobis distance as a function of hypothesised change index, k. Vertical lines indicate the beginning and end of the synthetic change transition.

predefined as $h = \lceil \frac{D+N+1}{2} \rceil$. This is the minimum allowable value of h for which the algorithm will still converge correctly. As such this method of estimation does not introduce any additional parameters. In its original form this method requires an exhaustive search over all possible subsets of size h which is prohibitive for large data sets. In [68] a modified version is presented which improves performance by orders of magnitude while converging to an identical or near identical result as [69]. This is the method that is applied in this work.

4.5.2 Gaussian Mixture Model

So far it has been assumed that all land cover patches exist on a spectrum between two end-members and all share the same natural variations. This is of course an oversimplification for the real case where there may also be spectrally very different land cover types. This introduces the need for a more expressive probabilistic model than a multi-variate Gaussian. Fortunately such a model need not be much more complex and can still be fit in an unsupervised manner. Another solution is to apply a clustering technique to estimate several models simultaneously such that each of them has a reduced variance. Gaussian Mixture Models (GMMs) are a fundamental unsupervised clustering model that have been successfully applied numerous times in the field of land cover classification and change detection [2]. They provide a simple way to extend the multi-variate Gaussian model above to the multi-modal case without requiring labeled training data. In a GMM the probability density is given by a linear weighted sum of Gaussian densities

$$\Pr(\mathbf{X}) = \sum_{i=1}^{C} \pi_i \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i). \tag{4.11}$$

The mean and variance for each component as well as the weights can be estimated in an unsupervised manner using the iterative Expectation maximisation (EM) algorithm. This introduces one additional parameter, C, which is the number of components. The choice of this parameter should ideally come from domain knowledge about the types of land cover in the region of interest, something that should not be difficult to estimate given a survey or visual inspection of satellite imagery. A purely analytical approach of minimising an information criteria could also be applied. It is important that the chosen value of C accurately reflects the number of distinct land cover classes in the region. Too

many modes may result in change/outlier time series being assigned their own clusters making them undetectable. Too few and the method retains the same problems of the empirical estimator.

If an accurate ground truth survey of the land cover is available this could be used to select the number of modes or a supervised approach similar to [43] may be used instead. This approach was not evaluated as the focus of this study is unsupervised methods for the case where ground truth is not available. It would however be a trivial modification to the proposed method, likely with improvements to detection performance.

Because of the additional modes introduced by using a GMM the methods proposed above require minor modifications. For the Mahalanobis detection method the distance is evaluated based on each of the modes and the smallest distance is thresholded. For the change point estimation method it is permissible that the time series may switch between modes at the change point. In this case the partitioned covariance matrix defined as

$$\mathbf{\Sigma}_{A\to B}(k) = \begin{bmatrix} \mathbf{\Sigma}_{A\ (1,\ldots,k^*)(1,\ldots,k^*)} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{B\ (k^*+1,\ldots,N)(k^*+1,\ldots,N)} \end{bmatrix}. \tag{4.12}$$

for the transition between components A and B. For each candidate change point the Mahalanobis distance is calculated for all permutations of A and B with repetitions and the smallest distance used to evaluate the change hypothesis. This introduces additional complexity as a covariance matrix must be estimated for each mode (linear in C) and the change point estimation method applied to all possible permutations of transitions between modes (polynomial in C). As discussed prior the number of modes should be small to prevent over-fitting and the increased complexity should not be prohibitive.

Both the change detection and change point estimation methods described above will be evaluated using the empirical, robust and GMM methods for estimating the joint density.

4.6 Experimental Design

In order to test the hypothesis a series of experiments is proposed that make use of both real and synthetic data. For a complete description of the data sets and including the selection of bands and rationale and design of the synthetic data see Chapter 6.

4.6.1 Performance Metrics for Offline Change Detection

Quantifying offline change detection performance requires a validation data set of D pixel time series that have been accurately classified as either experiencing change or no-change during the time period of interest. Given the decision output from a binary detection method it is possible to compute a 2×2 confusion matrix containing the counts of every possible outcome. These are the True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN)

$$\text{Confusion Matrix} = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix}. \tag{4.13}$$

From this matrix it is possible to derive a number of informative metrics for comparing detection methods. The most readily understandable of these is the *Accuracy Score*. This is defined as the fraction of examples classified correctly either as change or no-change

$$\text{Acc} = \frac{TP + TN}{D}. \tag{4.14}$$

While simple to understand this metric is misleading in detection problems that have a significant imbalance in the number of examples. As land cover change in most cases is a rare phenomenon (<5%), uniformly sampled validation data sets are comprised of majority no-change examples. In this case a detector that is 95% accurate could be formulated by declaring all pixels are no-change. As a result this metric is a poor choice for comparing methods or selecting parameters to maximise.

Far more acceptable in the field of detection theory is to evaluate both the True Positive Rate (TPR) and the False Positive Rate (FPR) defined as

$$\text{TPR} = \frac{TP}{TP + FN} \tag{4.15}$$

and

$$\text{FPR} = \frac{FP}{FP + TN}. \tag{4.16}$$

These give the fraction of positive examples classified correctly and the fraction of negative examples classified correctly respectively. In the absence of additional information

these metrics give an estimate of the probability that an example will be correctly classified given its true label is positive or negative respectively. An ideal method should simultaneously have a high TPR and a low FPR. It is easy to produce a method that performs well with regard to either TPR or FPR by either classifying all examples as either positive or negative. Getting the full picture requires considering both metrics simultaneously.

Methods that apply a threshold to discriminate between positive and negative examples present a trade-off between TPR and FPR. Assuming the change metric for all examples is finite it is possible to transition from all values classified as negative (TPR=0, FPR=0), to all examples classified as positive (TPR=1, FPR=1) by varying the threshold. By plotting the relationship between TPR and FPR parameterised by the threshold it is possible to produce a curve that describes the effectiveness of a given change metric, known as a Receiver Operator Characteristic (ROC) curve. This presents a method to compare detection methods that is invariance to selection of threshold. If a curve sits 'above' another curve it can be said that it is a superior detection metric.

A single comparison metric that is invariant to threshold and incorporates both the TPR and the FPR is calculated as the Area Under the ROC curve (AUROC). This is also equivalent to the Mann-Whitney U statistic which gives an estimate of the probability that a randomly selected positive example will have a detection metric higher than a randomly selected negative example [70]. One problem with this performance metric is it does not capture the sensitivity of the method to threshold selection which closely relates to the separability margin of the metric. Metrics with equivalent separability but different separability margins will produce an equal AUROC.

To compare proposed methods with methods from the literature the ROC curve and AUROC will be used. In addition an appropriate threshold will be selected and then used to compute the TPR and FPR. As previously mentioned the selection of threshold, or any parameter, adds significant difficulty when applying a method in practice as it requires obtaining a set of classified examples on which parameters are selected to optimise some performance metric. In the case of this experiment such a validation set already exits but to prevent an overestimate of performance on unseen data a final test set must be withheld. To get the best possible use of the data set a 5-fold cross-validation procedure is applied to estimate the threshold. This randomly splits the labeled data into 5 subsets.

All but one subset is used to identify a threshold which maximises some performance metric. This threshold is then used to classify the left-out subset and the performance in this set is recorded. This procedure is repeated leaving out a different subset each round. This results in 5 estimates of performance on unseen data. Another advantage of this approach is it allows for estimation of a confidence interval on the resulting performance metrics.

4.6.2 Performance Metrics for Change Point Estimation

Metrics for evaluating offline methods that identify the time-of-change are not so universally agreed upon. The review paper of Aminikhanghahi et al [54] includes a discussion of several different metrics with their advantages and disadvantages. For a data set containing D change points some candidate metrics include the Mean Squared Error (MSE)

$$\mathrm{MSE} = \frac{\sum_{i=1}^{D} \left(k_{\text{predicted}}^{(i)} - k_{\text{actual}}^{(i)} \right)^2}{D}, \tag{4.17}$$

Mean Absolute Error (MAE)

$$\mathrm{MAE} = \frac{\sum_{i=1}^{D} \left| k_{\text{predicted}}^{(i)} - k_{\text{actual}}^{(i)} \right|}{D}, \tag{4.18}$$

and Root Mean Squared Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}}. \tag{4.19}$$

Each of these methods is invariant to the sign of the error (e.g. predicted was before or after actual). The MSE, and hence RMSE, is particular sensitive to large errors making it unsuitable. For ease of interpretation the MAE is used to compare methods in this experiment. The histogram of errors will also be reported which can give much better insight regarding any outlying predictions or biases generated by a method.

4.6.3 Methods to Compare

As a comparison a simple maximum-likelihood, supervised off-line change detector was chosen. This method fits a 1-dimensional Gaussian density for each class at each time

step, yielding a mean and variance signal per class. It also makes the assumption that there are only two classes present, for example class A and class B, and the transition is always from A to B which is suitable for both the data sets considered. If the probability that an observation was generated by a A and B at time step i as $\Pr_i^{(A)}(x_i)$ and $\Pr_i^{(B)}(x_i)$ respectively, the maximum likelihood change point, k, can be defined as

$$\hat{k} = \arg\max_{1 \leq k \leq N} \prod_{i=1}^{k} \Pr_i^{(A)}(x_i) \prod_{i=k+1}^{N} \Pr_i^{(B)}(x_i). \tag{4.20}$$

This is valid as long as it is known that only transitions from A to B are present. There are also several publications [3], [6] that apply change detection and change point estimation methods to the same data set, settlement expansion in South Africa, allowing the results for that data set to be compared directly.

4.7 Conclusion

Based on the survey of the literature a possible new method for land cover change detection was motivated by the success of [59], [62] in using a Gaussian process to estimate the joint density. The proposed method instead models time series as a multivariate Gaussian PDF which has several advantages over a Gaussian Process in terms of prior parameter specification and computational complexity. Because it is fit directly from a collection of time series in a region it is also better able to incorporate information from neighbouring time series to estimate the region land cover dynamics, a key requirement for accurate and automated change detection.

In terms of user involvement in setting up these methods for use there is little that is required. The empirical and robust estimators require only a large, unlabeled training set of similar land cover for a region. This could be generated in a matter of minutes by selecting appropriate areas on a map that contain suitably similar land cover and are geographically close enough to experience similar vegetation phenology.

The output of the change detection methods is a metric, the Mahalanobis distance, which ideally increases with the probability of a time series containing a change point. For certain applications this may be enough and can be used to generate a map of possible change pixels. Other applications may require a discrete change/no-change flag and

therefore a suitable threshold must be selected. Once a decision has been made regarding the presence of change in a time series the time-of-change can be detected using the method described in this chapter while requiring no additional input.

These properties make the proposed methods excellent candidates for achieving the aims of total automation outlined in the previous chapters. The details and results of these experiments are contained in Chapters 6 and 7.

Chapter 5

Online Change Detection as Forecasting

5.1 Introduction

While the previous chapter concerned itself with identifying change points in fixed length time series, this chapter extends similar concepts to the online detection case. Online detection has very different applications and goals when compared with offline detection. Online land cover change detection is typically applied in monitoring applications where some action is to be taken upon the detection of change. It therefore follows that there are three main criteria by which a successful method must be evaluated. It must be able to detect changes with a high accuracy, low rate of false alarms and with minimal delay between the occurrence of a change and its detection. Much like the true/false positive rates, these metrics are at odds with one other. That is to say it is trivial to obtain a perfect score in one of them, while maintaining a suitable balance is a challenging problem.

The method of chapter 4 of using a collection of time series to model the land cover dynamics of a region can be readily applied to the online case through the realization that it is also possible to use a joint PDF to make predictions of the next value of a time series given its historical values. The difference between predicted and observed values gives a good indication if a time series is behaving as expected, given its history and its spatial neighbours, or if the observations are unexpected signifying a land cover change. This leads to the hypothesis for this chapter:

> *A time series of differences between observations and predictions made using a multivariate joint PDF fitted from a collection of land cover time series in a geographical*

region is uncorrelated, zero-mean Gaussian under no-change conditions and biased when the time series experiences a change point. This can be used to rapidly detect the occurrence of change.

Such a derived time series meets the assumptions required to apply classical statistical change detection algorithms such as the CUmulative SUM (CUSUM) algorithm [57], [58] or control charts [58].

The chapter begins with a discussion of statistical stationarity in signals and why this property is important for change detection. It is then shown how online change detection can be considered equivalent to forecasting with a sliding window and how a multi-variate joint PDF is well suited to this task. This expanded in to the description of a proposed method for online change detection which contains only the minimum number of free parameters and thus makes progress toward the goal of total automation.

This chapter also contributes a new method for performance evaluation of online change detection methods. The Detection Delay (DD) and Run Length to False Alarm (RLFA) are well known metrics for performance evaluation of online detection methods. This chapter presents a method for using finite length signals to estimate the survival function over these metrics for a given method applied to a particular data set. This enables statistical inferences to be made when comparing two or more methods. It is also shown how plotting the median DD vs the median RLFA is an effective way to compare methods and can be considered as the online detection equivalent to a ROC curve.

The content of this chapter has been compiled into a publication that appears in IEEE Geoscience and Remote Sensing Letters. See Appendix E or [12].

5.2 Background

5.2.1 Stationarity of Land Cover Time Series

A stochastic process can be said to be *strictly stationary* if the joint density of any combination of values is time invariant. For a strictly stationary process, $\{X_t\}$, it can be said for any set of time indices $\{t_1, \ldots, t_k\}$, the joint PDF

$$\Pr(X_{t_1+\tau}, \ldots X_{t_k+\tau}) \tag{5.1}$$

is identical regardless of the time shift τ. A special case of stationary is IID. As the name suggests each sample is independently drawn from the same univariate density. In this case the joint PDF may be factorised as

$$\Pr(X_{t_1+\tau}, \dots X_{t_N+\tau}) = \prod_{i=1}^{N} \Pr(X_{t_i+\tau}). \tag{5.2}$$

A process is said to be *cyclo-stationary* if the process obtained by sampling at integer multiples of some period T exhibits stationarity. Formally

$$\Pr(X_{t_1+T+\tau}, \dots X_{t_1+nT+\tau}), n \in \mathcal{Z} \tag{5.3}$$

remains identical for any value of t_1 and τ and for some period T.

Several factors combine to make remotely sensed land cover time series fail to meet these stationarity assumptions. Most land cover is vegetated and this results in the dominant source of non-stationarity in the signal. It is well documented that the reflectance of vegetation across most bands varies approximately periodically in a yearly cycle [71]. In many cases this could be assumed to be cyclo-stationary however the magnitude and bias of the oscillations is often non-constant between years due to inter-annual climatic changes such as droughts. In many cases trend components are present in vegetated time series that suggest either long term non-periodic change or change on a cycle much longer than the time span of available data.

The majority of online change detection algorithms perform best (and optimally in certain cases) when the time series is stationary under no-change conditions. Therefore each of online methods for land cover change detection reviewed in chapter 3 can be understood as being comprised of a signal transformation phase and a change detection phase (Table 5.1). The signal transformation phase is generally specific to land cover time series while the change detection algorithms are general purpose methods that can be applied to any stationary signal.

5.3 Change Detection by Forecasting

Consider an estimator that is able to produce forecasts of the next value in a time series under no-change conditions. The forecast at time t is denoted as $\hat{x}_t$. The residual of

TABLE 5.1: Tabulation of transformations and detection algorithms for online land cover change detection methods in the literature.

Publication	Signal Transformation	Change Detection Algorithm
[55]	Sliding window of parametric model parameters	Threshold
[9]	Sliding window of parametric model parameters	6 of 10 threshold heuristic
[56]	Kalman filter parameter series	CUSUM [57]
[59]	Gaussian Process prediction error	EWMA control chart [58]
[44]	Parametric model prediction error	MOSUM [49]
[4]	Z-score under supervised estimated density	CUSUM [57]

the forecast is defined as $r_t = x_t - \hat{x}_t$. A sufficiently good estimator should yield a series of residuals which are uncorrelated, unbiased (i.e. zero mean) and Gaussian. If the estimated residuals also have constant variance it can be said that the resulting time series of residuals is IID. Furthermore, if the residuals do not have constant variance but the estimator is able to produce good estimates of the variance of each sample, σ_t^2, the residuals can be converted to z-scores by

$$z_t = \frac{r_t}{\sigma_t} \tag{5.4}$$

and the series of z-scores will be IID unit Gaussian.

If the original time series contains a change point that is unforeseen to the estimator this will induce a bias in the series of z-scores/residuals which can be detected by a classical change detection method. An ideal estimator should yield a series of z-scores that is Gaussian IID under normal circumstances and respond strongly and rapidly in the presence of a change point. Such an estimator may take advantage of historical data, neighbouring pixel time series or ancillary data in order to make its prediction.

One such method that follows this general framework is the GPChange method of [59], [62]. In this case the prediction is made using only historical data from the same signal

$$\Pr(X_t) = \Pr(X_t \mid X_{t-1}, X_{t-2}, \dots X_0). \tag{5.5}$$

Motivated by these previous works a new method is proposed that takes advantage of both the previous samples in the time series and also the time series in the surrounding region when making a prediction. This is done by directly estimating the joint PDF over a temporal window using a set of similar time series. As the model is re-estimated at each time step this does not require any assumptions of stationarity. Furthermore it does

not impose assumptions on the structure of the data as in [9], [44], [72].

5.3.1 Proposed Method

As in the previous chapter it is proposed to model the joint PDF of all the samples in a fixed length time series as a multivariate Gaussian. In the online case however it must be over the samples in a fixed length temporal look-back window. Despite this limitation such a model is still able to capture the mean signal of the region of interest, the uncertainty associated with the mean signal and the correlations between samples in the time series. This statistical information can be leveraged to immediatly detect when a time series moves in a way that is inconsistent with both its own historical trajectory and that of the surrounding region.

With access to a fitted joint PDF it is possible to condition on the previous observations in a signal to obtain a univariate PDF over the last sample. For a multi-variate Gaussian PDF this can be computed efficiently and in closed form. More explicitly let $\boldsymbol{x} = [x_t, \ldots, x_{t-W-1}]^T$ be a vector of time series observations in a temporal window of size W. Let $\mathcal{X} = \boldsymbol{x}^{(i)}{}_{i=1}^{i=D}$ be a set of D time series from a similar geographical region and ideally containing similar land cover. It can be assumed that each of these time series windows is an observation of a multivariate random variable $\boldsymbol{X} = [X_t, \ldots, X_{t-W-1}]^T \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Under the multivariate Gaussian assumption the mean vector and covariance matrix can be calculated as

$$\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}] \tag{5.6}$$

and

$$\boldsymbol{\Sigma} = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] \tag{5.7}$$

which can be estimated directly from the data.

For a single time series, $\mathbf{x}^{(i)}$ this model can be conditioned to calculate the univariate density over the last time sample given earlier observations within some temporal window, $\Pr(x_t^{(i)} \mid \mathbf{x}_{obs}^{(i)})$, using the standard method for conditioning a multivariate Gaussian. If the covariance matrix is partitioned into observed and unobserved variables as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_t^2 & \boldsymbol{\sigma}_{1,2} \\ \boldsymbol{\sigma}_{2,1} & \boldsymbol{\Sigma}_{2,2} \end{bmatrix} \tag{5.8}$$

where

$$\boldsymbol{\sigma}_{1,2} = \boldsymbol{\sigma}_{2,1}^T = [\mathrm{cov}(X_t, X_{t-1}), \ldots, \mathrm{cov}(X_t, X_{t-W-1})] \tag{5.9}$$

and

$$\boldsymbol{\Sigma}_{2,2} = \begin{bmatrix} \sigma_{t-1}^2 & \ldots & \mathrm{cov}(X_{t-1}, X_{t-W-1}) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(X_{t-W-1}, X_{t-1}) & \ldots & \sigma_{t-W-1}^2 \end{bmatrix} \tag{5.10}$$

the conditional PDF is univariate Gaussian with mean and variance given by

$$\bar{\mu}_t^{(i)} = \mu_t + \boldsymbol{\sigma}_{1,2}\boldsymbol{\Sigma}_{2,2}^{-1}(\mathbf{x}_{\mathrm{obs}}^{(i)} - \mu_{\mathrm{obs}}) \tag{5.11}$$

and

$$\bar{\sigma}_t^2 = \sigma_t^2 - \boldsymbol{\sigma}_{1,2}\boldsymbol{\Sigma}_{2,2}^{-1}\boldsymbol{\sigma}_{2,1}. \tag{5.12}$$

From this probabilistic forecast the z-score of the observation of the i^{th} time series can be calculated as

$$z_t^{(i)} = \frac{x_t^{(i)} - \bar{\mu}_t^{(i)}}{\bar{\sigma}_t^{(i)}}. \tag{5.13}$$

Repeatedly applying this method as the sliding window is shifted results in a time series of z-scores. The time series of z-scores can then be monitored for persistent deviation from the standard normal which implies a change.

Again, as in Chapter 4, the empirical method for estimating the covariance matrix may not be the most suitable if the training set contains outliers or diverse types of land cover. Again the methods of rubust estimation or GMM can be trivially substituted. In the GMM case the the best fitting mode for the look-back window is used when making the forecast.

5.3.2 Change Detection on Z-Scores

Given a time series of z-scores produced by one of the methods discussed above it can be monitored to detect deviations from its expected distribution.To detect statistically significant persistent deviations from this distribution Page's CUmulative SUM (CUSUM) algorithm was chosen [57], [58]. The CUSUM method is presented under the assumption that the data is IID Gaussian with known mean and variance prior to change, after which it experiences a change in mean of unknown magnitude. This was selected over other

popular continuous monitoring algorithms, such as the Exponentially Weighted Moving Average (EWMA) control chart, due to its superior theoretical underpinnings and flexibility. According to [58] the two should be roughly equivalent in practice and have the same number of parameters.

Let $\{z_i\}_{i=0}^{i=N}$ be the series of z-scores. By definition this series should be IID Gaussian with zero mean and unit variance. It is therefore expected under normal conditions that the cumulative sum

$$C_i = \sum_{j=1}^{i} z_i \tag{5.14}$$

$$= z_i + C_{i-1} \tag{5.15}$$

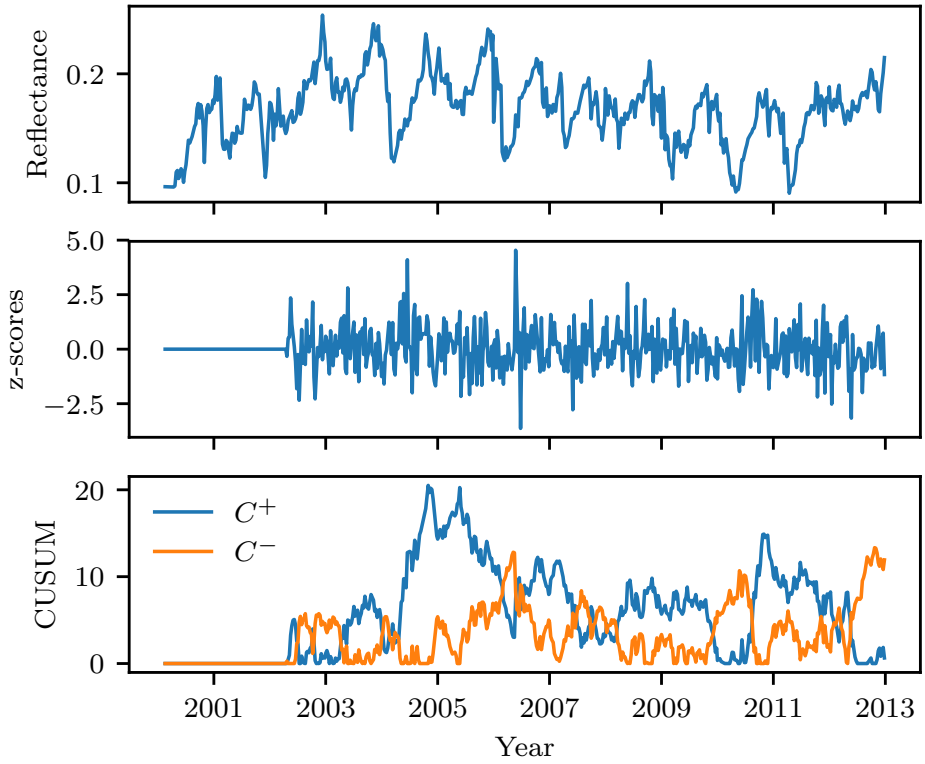
is a random walk with zero mean [58]. If there is a persistent bias in the models ability to make predictions, such as the underlying data generating distribution changing, the random walk is expected to drift away from zero in the direction of the bias. To improve the robustness of the method a slack parameter, k, is included [58]. This parameter is subtracted from the CUSUM at each iteration. It reduces the sensitivity to small mean shifts which can help prevent false alarms and also ensures the CUSUM decays back to zero if only z-scores smaller than k are observed. The cumulative sums are also split into positive and negative components with each clamped at zero. The resulting cumulative sums are defined recursively as

$$C_i^+ = \max[0, z_i - k + C_{i-1}^+] \tag{5.16}$$

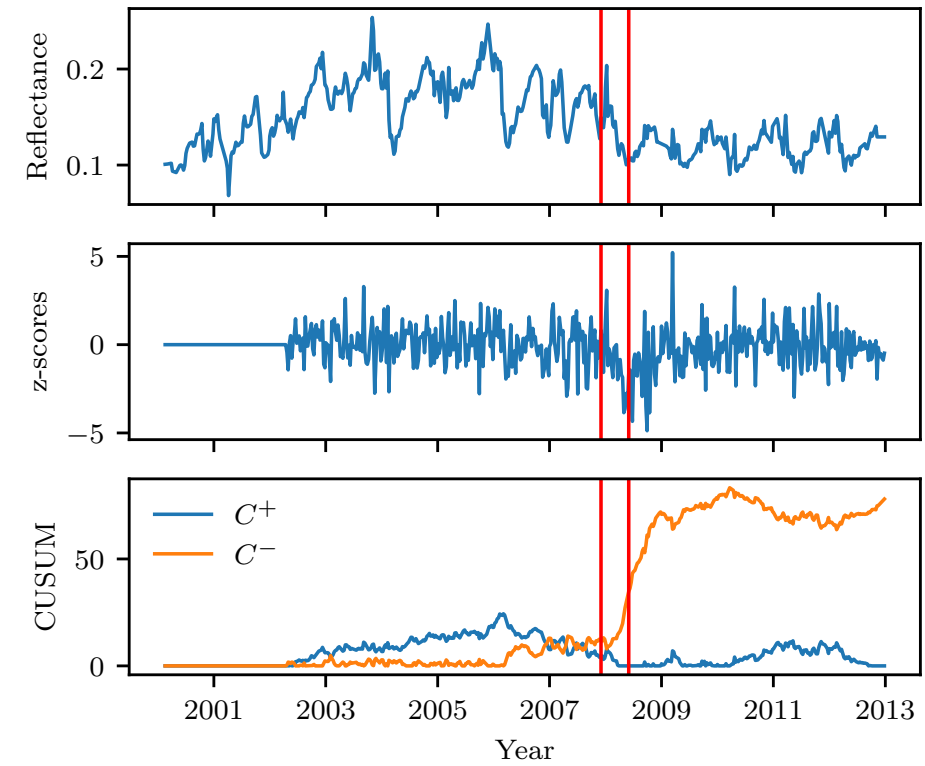
$$C_i^- = \max[0, -z_i - k + C_{i-1}^-]. \tag{5.17}$$

[58]

As it is not known if land cover change will cause a positive or negative change in the signal both cumulative sums are monitored. When either C^+ or C^- exceeds some threshold λ a change alarm is triggered and the corresponding cumulative sum is reset to zero. Fig. 5.1 shows a reflectance time series, z-score series extracted using the joint estimation method and the corresponding CUSUMs. In the time series containing a change it is possible to observe a negative bias in the z-score series immediately after the change



(A) No change time series



(B) Change time series

FIGURE 5.1: Reflectance time series, z-score series and CUSUM output for a no-change series and a synthetic change from the Limpopo data set. Vertical bars indicate the beginning and end of the linear transition.

point. This causes a strong response in C^-. In the no-change series the corresponding z-score series should be close to IID Gaussian noise and the CUSUMs of this signal will not significantly deviate from zero.

It can be shown that if the mean and variance before and after the change point is known and the change takes place instantaneously this method is optimal for a given detection delay [73]. This does not hold in this case as the distribution of the z-scores post change cannot be known. Furthermore the inclusion of a non-zero slack parameter breaks the optimality requirements for CUSUM [73].

5.4 Experimental Design

5.4.1 Performance Metrics for Online Change Detection

True/False positive rates give the probability of correct/incorrect detection at any time within the length of the signal. For continuous online detection this does not give any indication on the expected delay until detection. Instead we prefer to compare methods in terms of expected run lengths (i.e. time between alarms). This can be further split into two quantities, the Run Length to False Alarm (RLFA) and the run length to true alarm or Detection Delay (DD). The RLFA is defined as the number of samples after the algorithm starts, either at the beginning of the signal or after reseting due to a previous alarm, to the next alarm. The detection delay is defined as the number of samples between the actual change point and the next alarm. A good detection method strives to maximise the expected RLFA while minimising the expected DD [58]. These can be considered as the generalisation of false alarm rate and true alarm rate to the continuous monitoring case.

Estimating the distributions of RLFA and DD for a given method using finite length signals presents some challenges. As the z-scores signals are noisy the CUSUM method will always trigger an alarm if it is allowed to run for a sufficiently long time [58]. When estimating run length distributions from finite length signals this duration may be longer than the length of the signal. This can be taken into account by using the concept of censoring from survival analysis. The observation of no alarm gives information that the detection delay was at least as long as the remainder of the signal, known as right censoring. Failure to take this into account, for example by dropping series with no alarm,

results in an underestimate of run lengths. To account for the censoring of data we estimate the median detection delay by fitting a Kaplan-Meier estimator [74] a commonly used non-parametric method for estimating a survival function in the presence of censored data.

A survival function was estimated for both the RLFA and DD for each method at a series of thresholds. For RLFA if no alarm was triggered the run length was said to be at least the length of the signal. The RLFA was also considered to be censored when an actual change point occurred. For DD if no alarm was triggered after the change point the run length was said to be at least the difference between the change point and the end of the signal.

Plotting the median DD vs median RLFA with changing threshold yields a curve can be considered as a continuous inspection equivalent to a ROC curve. This will be referred to as a *run length curve*. Much like a ROC curve it allows for methods to be compared irrespective of threshold selection. Confidence intervals from the DD and RLFA estimates may also be included. One method can be said to be superior to another if its curve sits below another which can be interpreted as the method yielding a shorter detection delay for a given time between false alarms across the range considered.

5.4.2 Methods to Compare

Two methods in the literature stand out as suitable comparisons. The first is the supervised univariate density estimation method similar to that of [4]. The supervised univariate approach assumes knowledge of the land cover class of all time series in the training set as well as the initial class prior to change. A univariate Gaussian density is fit for each class at each time step. At test time z-scores are calculated using the appropriate density at each time step assuming the class does not change from its initial assignment.

The other method is the parametric forecasting method of [44]. This fits a parametric model with a bias, seasonal oscillation component and two higher harmonics over a fixed length look-back window. The model is refit at each time step as the window is shifted. The parametric model is used to make a forecast and the different between observation and prediction is converted to a z-score by dividing by the standard deviation estimated from the model residuals over the window. In the original paper this makes use of an

extended monitoring period during which the parameters of a seasonal harmonic parametric model is fit. A fixed length window was chosen instead such that the amount of historical data the model has access to can be controlled as part of the experiment.

Another modification made to the parametric method from its original form is the change detection method applied to the extracted series of z-scores. As shown in Table 5.1 this method originally made use of EWMA control charts for the continuous monitoring. To ensure a fair comparison between the other methods this was substituted with the CUSUM monitoring algorithm. Using the same algorithm for monitoring the z-scores ensures that any performance differences are due to the signal transformation stage only.

5.5 Conclusion

The method of estimating a joint PDF over samples translates well to the online case with the only modification to the estimation stage being the inclusion of a fixed length window. The benefits of requiring no labeled training set and no parameters to fit apply to this method equally as well. The way the fit PDF is used for detection however is very different. It would be possible to simply apply the same methods as Chapter 4 in a sliding window to produce an online detector but this would have severe limitations in terms of detection delay. This is because sufficient samples before and after the change point are required to make an accurate detection using the Mahalanobis distance method. This limitation is shared by [9] and [55] for these exact reasons and this is reflected in the reported detection delays.

By instead using the joint PDF to make forecasts and comparing these to the observations the detection is focused on the leading edge of the window. The increased sensitivity is mitigated by the CUSUM detector which introduces memory allowing the detector to become increasingly certain as more samples arrive supporting the change hypothesis until a given threshold is exceeded. Both the threshold and slack parameters of the CUSUM detector allow the operator to trade-off between detection delay (DD) and expected run-length between false alarms (RLFA) such that they are suitable for the particular application.

The select of these parameters is an important consideration and this chapter proposed a way to better estimate performance metrics such as DD and RLFA to assist with

this. The run length curve is much like a ROC curve for the continuous inspection case. Developed by estimating median DD and RLFA over a range of thresholds it similarly allows methods to be compared without threshold selection to identify the expected performance. It does however still require the selection of the one remaining parameter, the slack k, and sufficient examples of land cover change with which to estimate the survival function for DD.

Chapter 6

Study Areas and Results

6.1 Introduction

The previous two chapters introduced several hypotheses surrounding ways to improve the performance and levels of automation for detection of land cover change in both offline and online settings. A new method for change detection, change point estimation and online change detection was proposed based on the concept of leveraging a large number of time series in a region in order to estimate land cover dynamics, deviations from which constitute change.

Testing these hypotheses in a meaningful way requires evaluation on suitable problems and data sets that are representative of what might be encountered in a production system. Two problems are considered that assess two different types of land cover change, between class and within class. The first is the problem of detecting settlement expansion in rural South Africa. This particular problem was chosen as it has been the target of many research projects over the years making it an ideal testing ground to establish performance relative to other methods in the literature. The data set containing the land cover types and presence of change for a collection of MODIS pixels in the Limpopo province of South Africa was made available for use in this research project by CSIR, South Africa. The second problem involves the detection of deforestation or forest thinning in rural New South Wales, Australia. All of the locations used in this data set are currently under contract for preservation and generating carbon credits under the Emissions Reduction Fund (ERF) initiative. Effective automated detection of any changes in these regions has the potential to reduce the need for expensive auditing. This data set was created specifically for this project and has been made available to the research community via open access.

A simple method of generating a collection of synthetic time series is also discussed in this chapter. Such a data set has the advantage that all aspects of it can be controlled to isolate the influence of different parameters on the effectiveness of the methods.

This chapter begins by introducing in detail each of the data sets and steps taken to prepare the MODIS reflectance data. It also discusses the two ways in which synthetic data is created, the fully synthetic method and the hybrid synthetic change method. The remainder of the chapter consists of the results obtain by applying the proposed methods and the elected comparison methods to each of the problems. Additional experiments were conducted to assess the sensitivity of the methods to various parameters in order to better understand their workings and lend support to the hypotheses. The results will not be discussed in this chapter with a full discussion being reserved for Chapter 7.

6.2 Study Areas and Data

6.2.1 Study Areas

6.2.1.1 Settlement Expansion in South Africa

The first problem is that of detecting settlement expansion in the Limpopo Province of South Africa. The motivation for choosing this data set is several-fold. Primarily this problem has been studied extensively within the literature appearing in various forms in at least 13 publications by various authors [1]–[7], [45], [51]–[53], [75], [76]. Within the field of land cover change detection there are few problems that have been studied in as much detail as this one making it an ideal testing ground for new methods.

There are also some geographical and meteorological properties that make South Africa particularly well suited to applications making use of remotely sensed time series. Located at a latitude of around 23° S it has among the lowest annual cloud coverage of any land mass [77] making it an ideal candidate for applications of passive remote sensing data.

Finally, the detection of settlement expansion is an important problem for which there are measurable benefits from reducing the time to detection. In some areas of this province, settlement expansion is carried out ad-hoc by residents without prior planning

approval. Providing these settlements with utilities such as water, electricity and sewerage is the responsibility of the local governing body; however, the difficulty and cost associated with installing the infrastructure increases as the settlement matures. The local native vegetation of this region is predominately Savannah, characterised by small sparse trees and grassland [65]. Construction of settlements is characterised by the clearing of native vegetation and the construction of small dwellings [78].

A detailed description of the production of this data set can be found in [76]. To summarise, two high resolution images captured in the years 2000 and 2013 were visually overlaid and compared to identify areas which had been converted from vegetation to settlement within that time span. This change map was then overlaid with a MODIS pixel grid and any pixels covered at least 70% by change were labeled and added to the data set. No-change pixels were produced by an initial classification of polygons as either vegetation or settlement within the region. MODIS pixels from these polygons were randomly sampled if they did not contain any significant changes within the time span and added to the data set. The final data set contains $D = 2348$ MODIS pixel time series classified as *vegetation* - 997 (42%), *settlement* (53%) - 1235 or *change* - 116 (5%). Due to contractual reasons the coordinates of the MODIS pixels for this dataset cannot be published.

6.2.1.2 Deforestation in Australia

The second problem relates to detecting changes in areas of native forest in New South Wales, Australia. Under the Australian government Emissions Reduction Fund (ERF) initiative, land owners are able to generate Australian carbon credit units (ACCUs) in exchange for preventing deforestation of native forest for which a clearing permit has previously been issued. To be eligible the native forest must have at least 20% canopy coverage with tree height greater than two meters [80]. At the commencement of a project an extensive audit is undertaken by an accredited third party and estimates made of the appropriate number of credits to be allocated throughout the project. A number of subsequent audits are also required to ensure compliance throughout the project lifetime. Projects have a permanence period of either 25 or 100 years. During this time period no clearing, with the exception of minimal thinning ($< 5\%$), is permitted [80].

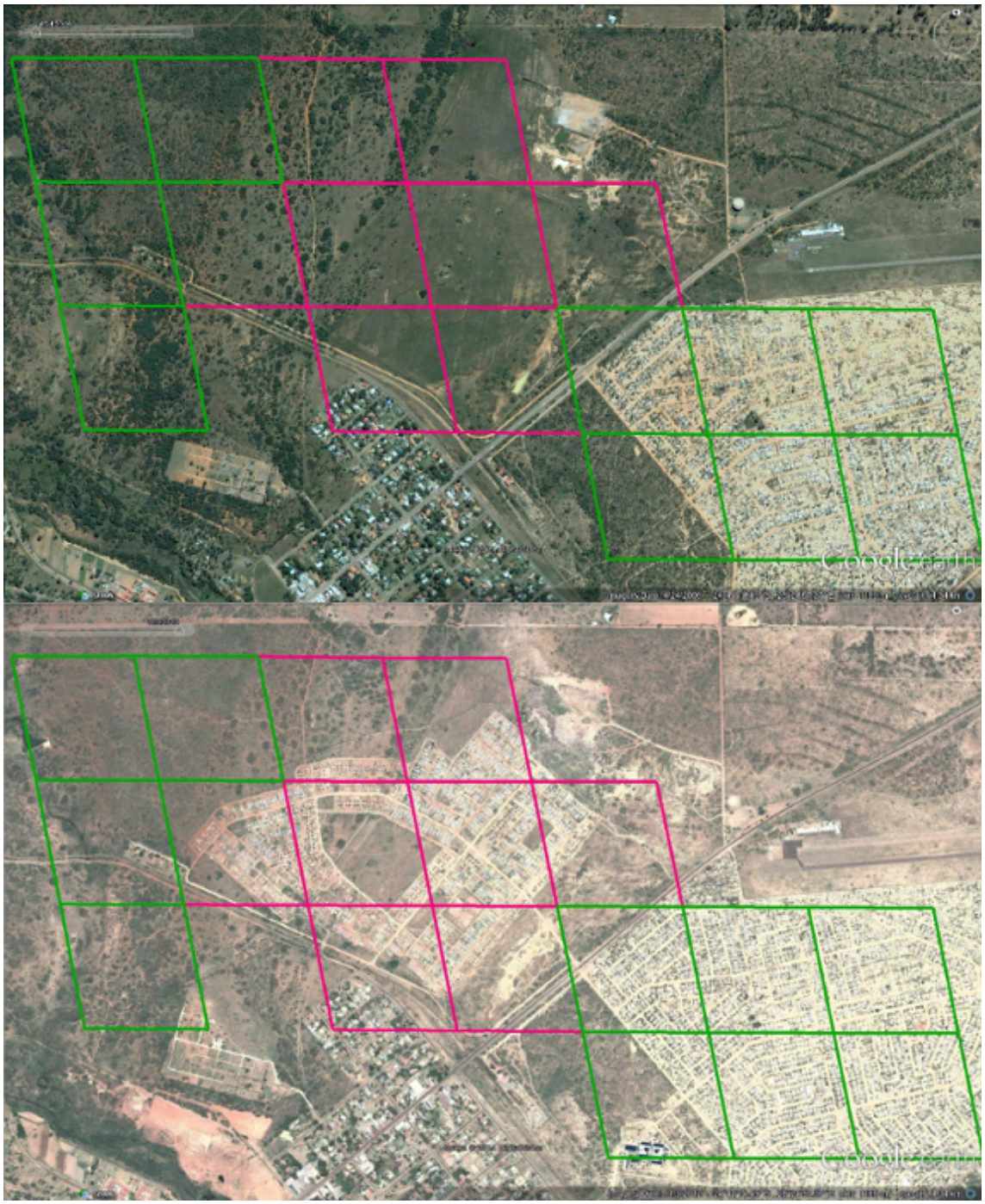


FIGURE 6.1: Example MODIS pixels (overlaid grid) of vegetation, settlement and pixels experiencing change in Limpopo Province, South Africa. (Figure originally appears in [79]. Reproduced with permission of the author)(Backdrop image courtesy of Google Earth, DigitalGlobe. 0.65m resolution optical data)

At the time of writing there are approximately 400 vegetation projects underway in Australia. A system of automated change alarms based on remotely sensed time series has the potential to significantly reduce the auditing workload and target it to locations where the forest cover has changed and the project may require reassessment. Continuous monitoring is also important to guarantee the integrity of the carbon credit units in circulation.The region considered occupies the western plains of the state of New South Wales. This is an arid region with the majority of land cover characterised by small shrubs. The forested regions protected under the ERF are predominantly free standing Eucalyptus trees [81]. The western plains also experiences very low cloud cover [77] making it another ideal location for passive remote sensing.

This data set was created by the author and is available online with the intention that it will also be the subject of future studies [82]. Due to the comparative scarcity of actual land cover change in this region, the data set was created for the purpose of synthesising realistic change time series. $D = 997$ MODIS pixels were identified within regions currently assigned to a project which is currently earning carbon credits in exchange for avoided deforestation. These time series span the years of 2008 to 2018. For each pixel a corresponding pair was identified using high resolution imagery provided by Google Earth, DigitalGlobe. The matching pair for each pixel was selected as a nearby pixel containing identical land cover but at a lower density. Fig. 6.2 illustrates an example of a pixel pairing. Using this approach it is possible to generate a synthetic change time series by blending the two while reducing any possible changes in geology, flora, aspect or climate that might occur over a large distance. Blending between high and low density aims to simulate gradual deforestation over the area of the pixel.

6.2.2 Data Preparation

For both applications the time series were obtained from the MODIS MCD43A4 product [24]. As discussed in Chapter 2 this product delivers Bi-Directional Reflectance Distribution Function (BRDF) corrected surface reflectance data from both the Terra and Aqua Platforms. The temporal resolution is 8 days with each data point selected as the highest quality acquisition in a 16 day temporal window. The spatial resolution is approximately 500m. This product has been used in numerous land cover change detection studies mostly due to its high temporal resolution and large catalogue of data. The Normalised

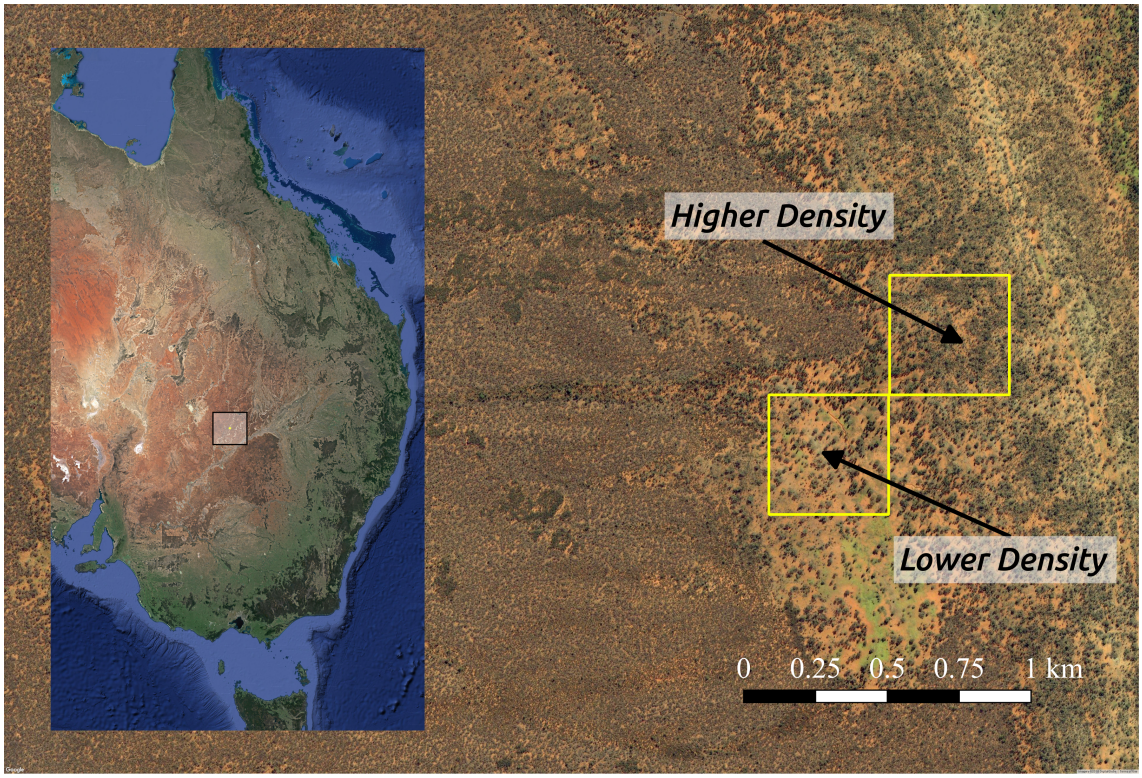


FIGURE 6.2: The study region in New South Wales. This includes an example of two MODIS pixels within a protected area that contain similar land cover with differing densities. This pair of pixel time series is blended together over a 6 month transition to simulate the partial and gradual clearing of vegetation. (Backdrop image courtesy of Google Earth, DigitalGlobe. 0.65m resolution optical data)

Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) were also considered and derived from the MCD43A4 product. Initially the data sets were screened for missing data due to cloud cover, sensor malfunction or saturation. Due to the lack of clouds in these regions coupled with the process of taking the best acquisition in the 16 day window, both data sets contained negligible missing values ($< 0.1\%$). The few missing values present were filled using linear interpolation.

6.2.3 Hybrid Synthetic Change Data

Producing data sets of real land cover change is a very challenging problem for several reasons. Firstly, for the study areas discussed above land cover change is a rare occurrence. Even if it were possible to correctly identify all of the regions that experience change within the time span of interest there may not be enough data to yield statistically significant results. Secondly, identifying the regions that have changed and verifying their correctness is difficult and time consuming. It is necessary either to obtain very high resolution optical imagery for two dates to manually compare (as for the Limpopo data set), to conduct on-ground surveys, or a combination of both. Both are expensive, time consuming and require special expertise.

Even after one of the above methods have been applied it is very difficult to identify the exact time span over which the land cover transition occurred. When manually inspecting imagery it is only possible to declare that a change occurred between two acquisition dates which are typically years apart. This is not sufficient for evaluating the performance of change point detection methods or online detection methods. As such almost all studies that assess the time-of-change make use of synthetic change data.

Kleynhans et al [6] describe a method for generating hybrid synthetic change time series has been applied in numerous other works [3], [8], [9], [51], [53], [55], [56], [83]. Change time series are generated from real no-change series by randomly selecting two that are known to contain different land cover, randomly selecting a change point from a uniform distribution and then linearly blending the two series over a suitable length window. Selection of window size is dependent on the nature of the change the system is designed to detect. Kleynhans et al found a window of 6 months to be suitable for detecting conversions from vegetation to human settlement and the same window size is used for both data sets in this study. The linear blending aims to simulate gradual

nature of land cover conversions over the large size of MODIS pixels. For higher spatial resolution data or for certain applications such as fire detection this may not be required. To produce hybrid synthetic change data sets from the MODIS time series, in this work the same method is applied with the additional constraint that change may not occur within a six month buffer of the beginning or end of a series.

6.2.4 Fully Synthetic Data

In order to have complete control over the statistical properties of the time series, experiments are also conducted on fully synthetic data. This makes it possible to conduct highly controlled experiments to evaluate how the methods perform with different levels of noise, change magnitudes, training set noise and signal length. The synthetic signals were designed to mimic certain properties of vegetated land cover time series such as seasonal variation and trending growth. The collections of time series generated also aim to be similar to collections of time series from a region by including random bias and trends. The synthetic time series were generated from the model

$$x_i[t] = A_i \cos\left(\frac{2\pi t}{T}\right) + B_i[t] + Z[t] \tag{6.1}$$

where Z is Additive White Gaussian Noise (AWGN) added to each sample in the time series to simulate noise in the sensor system

$$Z[t] \sim \mathcal{N}(0, \sigma^2_{\text{noise}}). \tag{6.2}$$

A_i is the amplitude for time series i. In this experiment this was kept constant

$$A_i = 1. \tag{6.3}$$

Similarly each time series is offset by a bias function which was defined as a linear function with random variables describing the slope and intercept

$$B_i[t] = \frac{M_i}{L}\left(t - \frac{L}{2}\right) + \alpha_i \tag{6.4}$$

$$M_i \sim \mathcal{N}(0, \sigma^2_{\text{slope}}) \tag{6.5}$$

$$\alpha_i \sim \mathcal{N}(0, \sigma^2_{\text{bias}}) \tag{6.6}$$

where L is the length of the time series.

This allows for simulating both noise in the time series themselves but also the variations that exist between time series in the data set, such as different offsets, which aim to simulate different pixel compositions; and slope, which simulates different growth rates. Changes are simulated through the addition of a step function to the bias. The location of the step is selected uniformly randomly across the length of the time series excluding samples within $\frac{T}{2}$ of the beginning and end. A change series is therefore given by

$$x_{i,change}[t] = x_i[t] + A_{\text{step}} H[t - k^*] \tag{6.7}$$

$$k^* \sim \mathcal{U}(\frac{T}{2}, L - \frac{T}{2}) \tag{6.8}$$

where A_{step} is the step size, L is the length of the time series and $H[]$ is the Heaviside step function. It is also useful to consider the step size in terms of the noise power, σ^2_{noise}, giving the Signal to Noise Ratio (SNR)

$$\text{SNR} = \frac{A^2_{\text{step}}}{\sigma^2_{\text{noise}}} \tag{6.9}$$

6.3 Results

6.3.1 Offline Detection

6.3.1.1 Change Detection

The proposed method for change detection using Mahalanobis distance and three different methods for estimating the joint PDF (*empirical*, *robust* and *GMM*) was applied to both real change data from the Limpopo data set and the hybrid synthetic change data for the NSW data set. Additionally the suggested comparison method of assuming uncorrelated

samples but estimating a per-class density using a supervised training procedure (*uncorrelated*), was also evaluated. In each case, excluding the supervised method, the joint PDF was fit from the entire data set including the change time series and 5-fold cross validation was used to both select and appropriate threshold and estimate the Accuracy (Acc), True Positive Rate (TPR), False Positive Rate (FPR) and Area Under ROC curve (AUROC) along with their standard deviations across the folds. For the supervised method, the same cross-validation procedure was applied but only the no-change time series in the training set were used to fit.

The above procedure was applied to all bands plus the derived indices NDVI and EVI. The table of results is very large and therefore only the table of best performing bands for each method by AUROC are included in text and given in Table 6.1 for Limpopo province and Table 6.2 for NSW. The complete set of results can be found in Appendix A Tables A.1 and A.2. The ROC curves were plotted, again for the best performing bands for each method, in Fig. 6.3a and Fig. 6.3b.

Table 6.3 gives a comparison of the best performing bands for each of the evaluated methods, with two other results published in the literature for the Limpopo province data set. These two studies were selected as they both give cross validated performance assessments and publish results of TPR, FPR and AUROC making meaningful comparisons possible. As estimates of uncertainty were not published in other papers they cannot be included for comparison.

TABLE 6.1: Test set results for offline change detection applied to real change data from the Limpopo settlement expansion set. Mean values are given in percentage with standard deviation across all cross validation folds given in parenthesis. Results are sorted descending by AUROC. Only the best performing bands by AUROC are included for each method. Full table is included in Appendix A

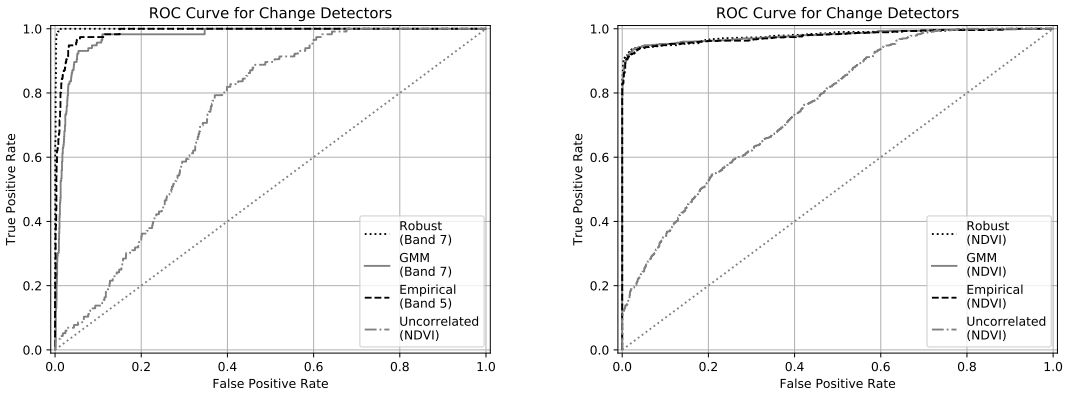
Band	Method	AUROC	TPR (%)	FPR (%)	Acc (%)
7	robust	0.999	93.91(12.18)	1.16(0.59)	98.59(0.37)
5	empirical	0.989	93.92(6.51)	4.93(1.69)	95.02(1.36)
7	GMM	0.974	95.69(6.73)	9.72(2.86)	90.55(2.57)
NDVI	uncorrelated	0.726	98.26(2.13)	65.29(20.09)	37.86(18.98)

TABLE 6.2: Test set results for offline change detection applied to the synthetic change data from the New South Wales devegetation set. Mean values are given in percentage with standard deviation across all cross validation folds given in parenthesis. Results are sorted descending by AUROC. Only the best performing bands by AUROC are included for each method. Full table is included in Appendix A

Band	Method	AUROC	TPR (%)	FPR (%)	Acc (%)
NDVI	robust	0.980	91.17(5.12)	1.40(0.86)	94.88(2.18)
NDVI	GMM	0.979	92.38(4.40)	2.31(0.75)	95.04(1.99)
NDVI	empirical	0.977	90.97(5.46)	1.91(0.97)	94.53(2.51)
NDVI	uncorrelated	0.757	95.09(5.04)	66.70(4.39)	64.19(1.13)

TABLE 6.3: Comparison of performance metrics for Limpopo province with other methods in the literature applied to the same data set. Mean values are given in percentage with standard deviation across all cross validation folds given in parenthesis. Bands with the highest TPR are shown.

Method	TPR (%)	FPR (%)	Acc (%)	AUROC
Robust Covariance (B7)	93.91(12.18)	1.16(0.59)	98.59(0.37)	0.999
GMM (B7)	95.69(6.73)	9.72(2.86)	90.55(2.57)	0.974
Empirical Covariance (B5)	93.92(6.51)	4.93(1.69)	95.02(1.36)	0.989
Extended Kalman Filter (EKF) [6]	89	13	87	-
EKF + Non-linear Detector (7 bands) [3]	96	0.2	99	-
EKF + Non-linear Detector (1 band) [3]	96	1	99	-



(A) ROC plot for the Limpopo data set.

(B) ROC plot for the NSW Data set.

FIGURE 6.3: Receiver Operating Characteristic (ROC) curves for the evaluated change detection methods. Only the best performing band according to TPR is shown.

6.3.1.2 Change Point Estimation

Since the exact change points were not known for the Limpopo data set, synthetic change time series were used for both sets to evaluate the change point estimation performance.

The same methods for estimating the joint PDF from the previous section were applied to detecting the time-of-change using the proposed method of partitioning the covariance matrix. In the case of the supervised uncorrelated model this had the added information about the land cover class before and after the change (vegetation to settlement). Again the methods were evaluated across all bands plus NDVI/EVI and the best performing bands by Mean Absolute Error (MAE) are included in Tables 6.4 and 6.5 while the full tables are available in Appendix A Tables A.3 and A.3 for Limpopo and NSW respectively. The full distribution of errors for both data sets on the best performing bands are included in Fig. 6.4a and Fig. 6.4b.

TABLE 6.4: Comparison of mean error (Mean) and Mean Absolute Error (MAE) between estimated and known change point on synthetic change data from the Limpopo data set. Standard deviation across all cross validation folds given in parenthesis. Errors are given as the number of samples. Only the best performing bands by AUROC are included for each method. Full table is included in Appendix A

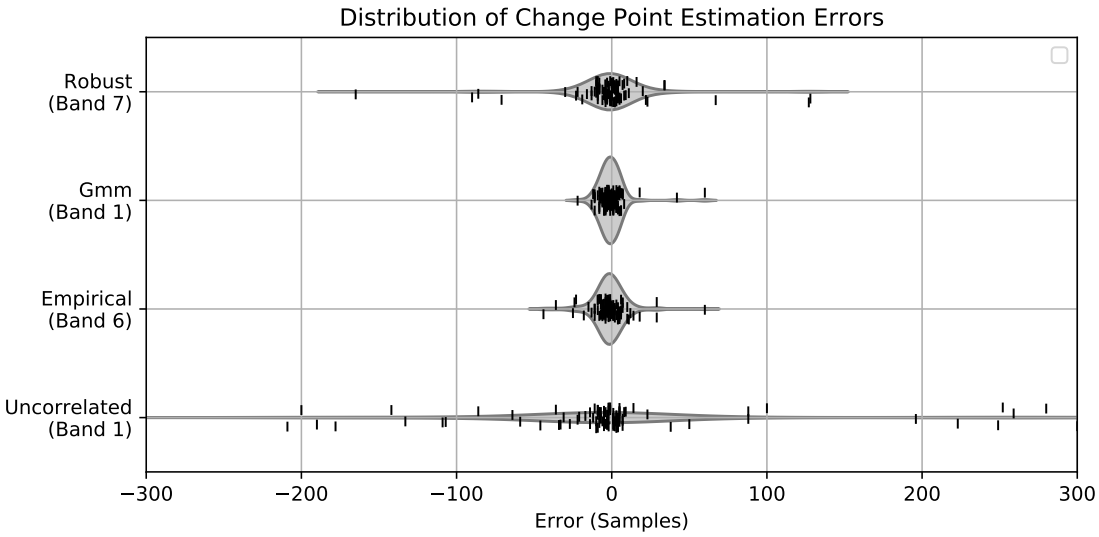
Band	Method	Mean	MAE
1	GMM	-0.26	5.38(7.57)
5	empirical	-0.26	6.92(10.82)
7	robust	-1.40	13.52(27.47)
1	uncorrelated	6.86	52.44(88.97)

TABLE 6.5: Comparison of mean error (Mean) and Mean Absolute Error (MAE) between estimated and known change point on synthetic change data from the NSW data set. Standard deviation across all cross validation folds given in parenthesis. Errors are given as the number of samples. Only the best performing bands by AUROC are included for each method. Full table is included in Appendix A

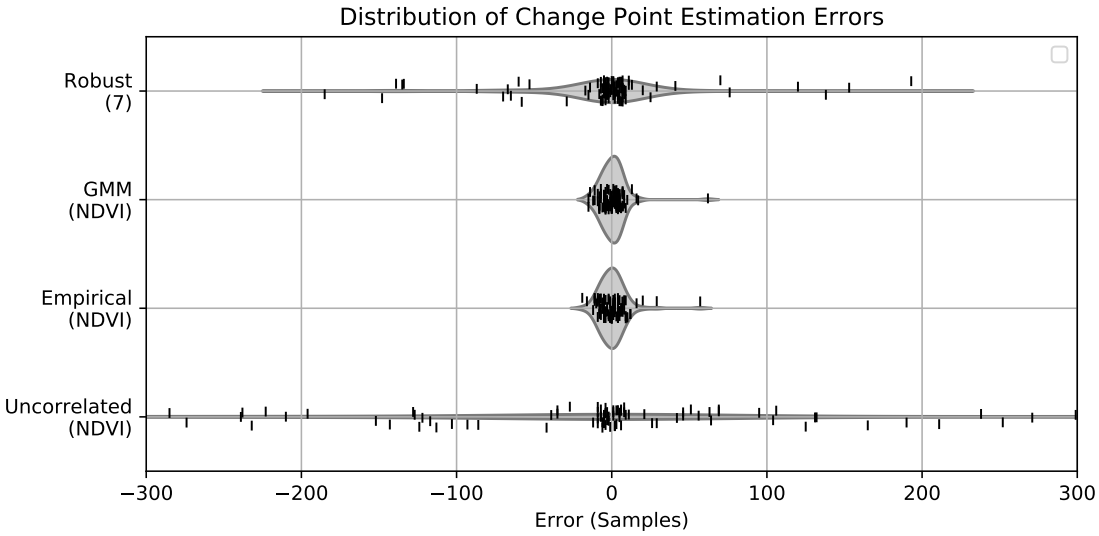
Band	Method	Mean	MAE
NDVI	GMM	0.62	5.52(6.82)
NDVI	empirical	0.62	6.00(6.98)
7	robust	-3.67	24.87(43.90)
NDVI	uncorrelated	-1.03	110.11(122.57)

6.3.1.3 Synthetic Data Experiments

In order to estimate how the method is influenced by the signal-to-noise ratio (SNR) of each time series as well as the level of signal diversity in the training set, several experiments were conducted on synthetic data generated using the method described above. The between signal variation was introduced by several different mechanisms, by adding



(A) Error distribution plot for the Limpopo data set.



(B) Error distribution plot for the NSW Data set.

FIGURE 6.4: Change point estimation error distributions for the evaluated change detection methods. Only the best performing band according to MAE is shown.

a Gaussian random variable to the offset and also to the slope of the bias. In order to encapsulate the between-signal variation in a single value the entropy of the estimated PDF

$$S = \frac{L}{2} + \frac{L}{2}\ln(2\pi) + \frac{1}{2}\ln(|\mathbf{\Sigma}|) \tag{6.10}$$

was calculated. Table 6.6 gives the different values of the entropy of the joint PDF estimated using the empirical covariance approach. These were calculated for a fixed noise power $\sigma_{\text{noise}} = 1.0$. Plots of the SNR vs AUROC and SNR vs MAE were produced for

various levels of entropy in Fig. 6.5.

TABLE 6.6: Entropy of the joint PDF estimated using from a synthetic data set using the empirical covariance method for different values of the bias and slope noise.

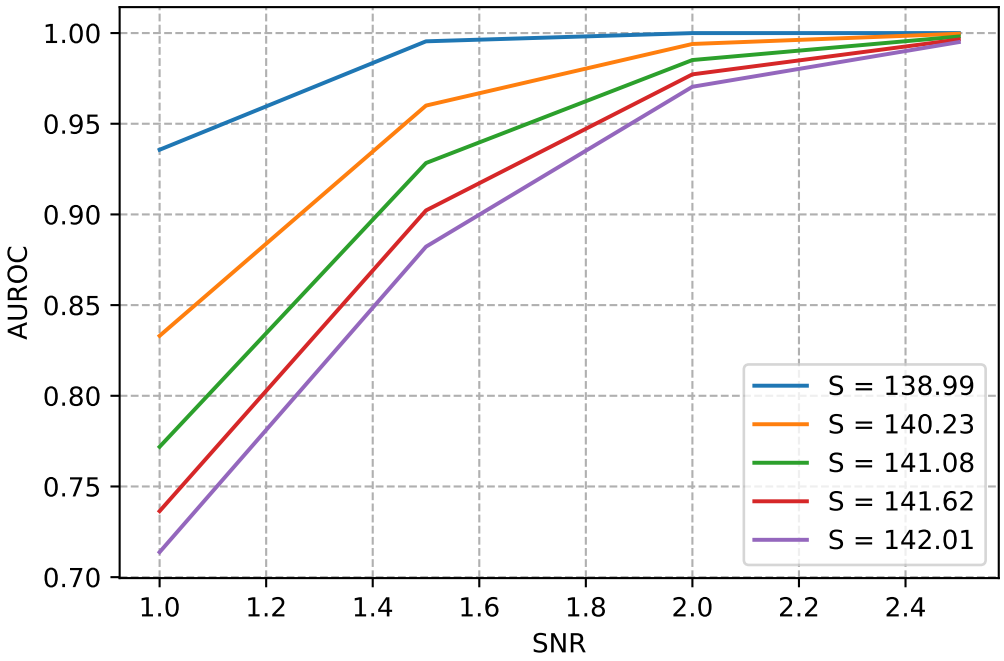
σ_{bias}	σ_{slope}	Entropy, S
0.0	0.0	138.99
0.06	0.01	140.23
0.11	0.02	141.08
0.14	0.03	141.62
0.16	0.04	142.01

6.3.2 Online Detection

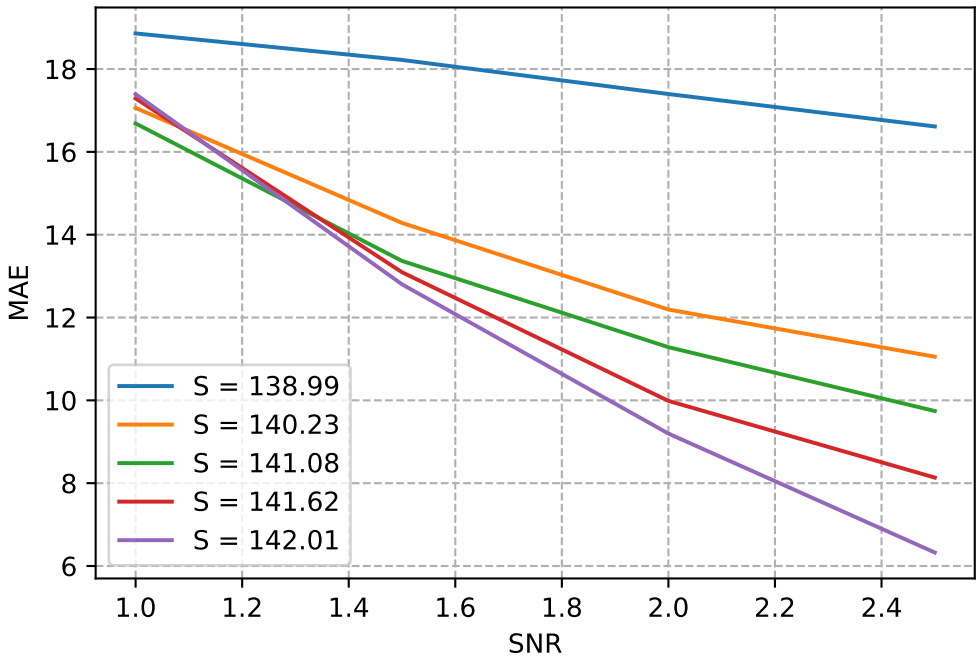
For the online detection case both data sets were used. In this case only the GMM method for estimating the joint PDF was evaulated and will be referred to as *GMM Joint*. The other methods implemented for comparison are the parametric model method of [44], *Parametric*, and the online detection variant of estimating class conditional univariate densities of [4], *Supervised Univariate*. Again since the exact time of change cannot be known only hybrid synthetic change time series are considered for both data sets.

The first series of experiments does not involve making predictions but rather addresses the first part of the hypothesis of chapter 5: *"The time series of differences between observations and predictions made using a multivariate joint PDF fitted from a collection of land cover time series in a geographical region is uncorrelated, zero-mean Gaussian under no-change conditions."*

This part itself can also be divided further into two assertions: the time series of differences is *uncorrelated* and it is *zero-mean Gaussian*. Fig. 6.6 shows the histogram and autocorrelation for a single change and real (non synthesised) no-change time series from the Limpopo data set. Recall that for a perfectly uncorrelated time series the autocorrelation should be a unit impulse centered at zero. While this figure provides good insight and suggests for a single time series that the derived series of z-scores has the desired properties, it is not sufficient to test the hypothesis. For this purpose the Kolmogorov–Smirnov test [84] is applied to test if the derived z-scores are normally distributed. Under this test the null hypothesis states that the samples are generated by the reference distribution (in this case unit Gaussian). Values of the D-statistic close to zero support the null



(A) AUROC as a function of SNR for various values of the entropy.



(B) MAE as a function of SNR for various values of the entropy.

FIGURE 6.5: The AUROC and MAE as a function of SNR for different values of between signal variation given by the entropy of the estimated PDF.

hypothesis. A p-value greater than the given alpha level ($\alpha = 0.05$ in this case) implies that the null could not be rejected and supports the hypothesis that the data is Gaussian distributed.

The Ljung-Box test [85] is applied to ascertain if the derived z-scores are uncorrelated. This test compares the autocorrelation function with the ideal uncorrelated case taking into account the number of samples used to estimate it. The null hypothesis is that the series is uncorrelated, which is supported by large values of the Q-statistic. Again a p-value greater than the given alpha level implies that the null could not be rejected thus supporting the hypothesis that the data is uncorrelated. For this experiment the autocorrelation with lags up to 10 samples was considered.

TABLE 6.7: Table of test statistics and p-values for the Kolmogorov–Smirnov (K-S) and Ljung-Box (L-B) tests applied to the derived z-scores for the Limpopo data set. Values in boldface indicate cases where the null could not be rejected. Only the first two bands are included here. Full table is included in Appendix A

method	band	D	p (K-S)	Q	p (L-B)
GMM Joint	1	0.075	0.008	9.5	**0.48**
GMM Joint	2	0.037	**0.5**	1e+01	**0.41**
Unsupervised	1	0.25	0.0	5.4e+02	5.3e-109
Unsupervised	2	0.36	0.0	1.6e+02	3.4e-29
Parametric	1	0.099	0.00012	7.3e+02	8.8e-150
Parametric	2	0.058	**0.066**	3.8e+02	3e-76

TABLE 6.8: Table of test statistics and p-values for the Kolmogorov–Smirnov (K-S) and Ljung-Box (L-B) tests applied to the derived z-scores for the NSW data set. Values in boldface indicate cases where the null could not be rejected. Only the first two bands are included here. Full table is included in Appendix A

method	band	D	p (K-S)	Q	p (L-B)
GMM Joint	1	0.032	**0.89**	3.7	**0.96**
GMM Joint	2	0.091	0.0089	1.2e+01	**0.28**
Unsupervised	1	0.81	0.0	9.6e+02	4.2e-200
Unsupervised	2	0.59	0.0	1.6e+03	0.0
Parametric	1	0.18	1.5e-09	1.7e+03	0.0
Parametric	2	0.22	0.0	1.5e+03	0.0

6.3.2.1 Change Detection Assessment

The performance of the online change detection methods was first considered in terms of detection ability only, without considering delay. This allowed the method to be compared with offline detection methods. All methods require selection of the slack parameter, k, and both the parametric and the joint Gaussian estimation methods require a finite length look-back window of size W in order to make a forecast. The ROC curve was evaluated for a window size of 100 and with the best slack parameter selected from the curves in Fig. 6.7. Five-fold cross validation was used to estimate the mean ROC curve and its 95% confidence interval. For the Limpopo data set, real change data was used while for the NSW data set synthetic change series were combined with real no-change time series. Fig. 6.8a gives the curves for the Limpopo data set and Fig. 6.8c gives the curves for the NSW data set.

To visualise the influence of the window size and slack parameters Fig. 6.7 gives the Area Under ROC curve (AUROC) for detection of change/no-change time series as a function of k for several window sizes. The AUROC and TPR at FPR=0.2 are tabulated for each of the methods with the best performing parameter configurations in Tables 6.9 and 6.10.

6.3.3 Detection Delay Assessment

Quantifying performance of detection delay in the online case is less trivial than for offline detection since one must account for the possibility of several alarms being triggered in the same time series and the fact that alarms may not be observed within the duration of the finite length time series on which the experiment was conducted. This makes estimating the expected detection delay (DD) and time between false alarms (RLFA) a challenging problem requiring its own probabilistic model. A survival function was estimated for both the RLFA and DD for each method at a series of thresholds using a Kaplan-Meier estimator [74], a method for estimating survival functions in the presence of censored data. The survival curve gives the probability that no alarm has been seen as a function of the number of samples since the previous alarm, for the RLFA case, or since the actual change point for the DD case. For RLFA if no alarm was triggered the run length was said to be censored at the length of the signal. The RLFA was also considered

to be censored when an actual change point occurred. For DD if no alarm was triggered after the change point the run length was said to be at least the difference between the change point and the end of the signal.

The run length curves were produced by evaluating the median of the DD and RLFA survival curves at multiple threshold values. The DD and RLFA are then plotted against each other similar to a ROC curve. The curves also include an estimate of uncertainty in the form of the 95% confidence interval of the median of the DD for a given RLFA. Fig. 6.8b gives the run length curves for the Limpopo data set and Fig. 6.8d gives the curves for the NSW data set for MODIS band 1 for each of the methods considered. The ROC and Detection Delay plots for other bands are included in Appendix B. In some cases there were not enough true alarms to successfully fit the Kaplan-Meier estimator for DD. In these cases the run length curves are omitted. The median DD for an arbitrarily picked median RLFA of 200 samples (1600 days at the MODIS sample rate) is reported in Tables 6.9 and 6.10.

TABLE 6.9: Results for detection of settlement expansion in Limpopo Province. Only the best performing bands by AUROC are included for each method. Full table is included in Appendix A.

Method	AUROC	Band	TPR at FPR=0.2	DD at RLFA=200
Parametric	0.86	4	**87.03(3.97)**	22.00
GMM Joint	0.87	7	75.94(6.65)	**10.00**
Supervised	**0.90**	1	86.20(3.29)	28.00

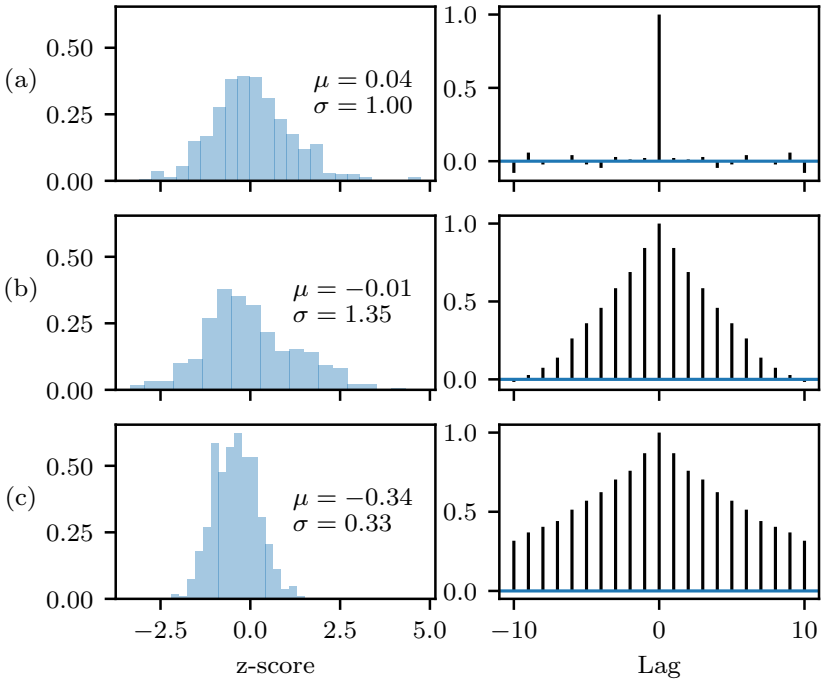
TABLE 6.10: Results for detection of vegetation thinning in New South Wales. Only the best performing bands by AUROC are included for each method. Full table is included in Appendix A

Method	AUROC	Band	TPR at FPR=0.2	DD at RLFA=200
Parametric	0.76	6	54.76(5.24)	64.00
Supervised	0.68	2	46.34(2.69)	315.00
GMM Joint	**0.84**	4	**75.64(2.64)**	**13.00**

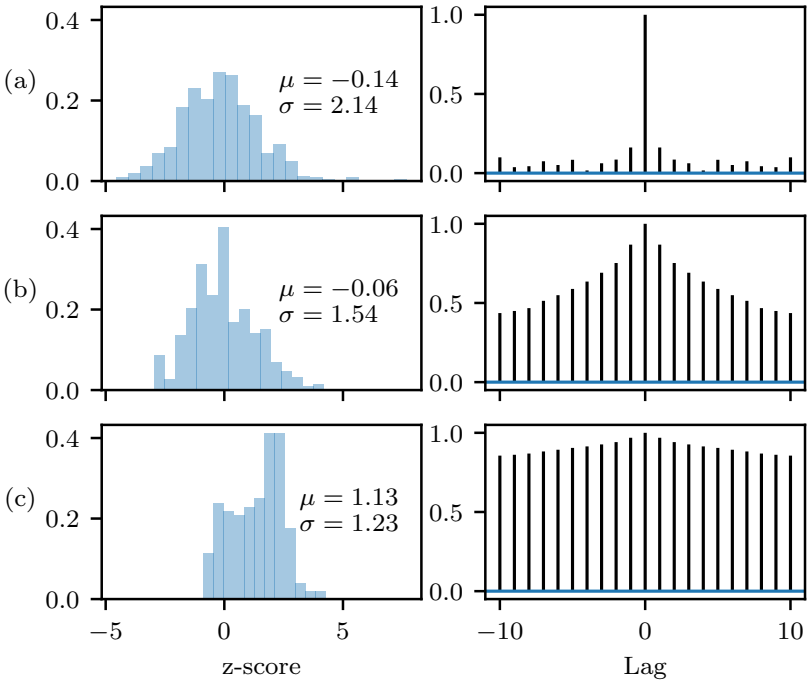
6.4 Conclusion

This chapter presents the reader with the results required to evaluate the hypotheses stated in chapters 4 and 5 including the relative performance of proposed and comparison method for change detection, change point estimation and online change detection,

and the statistical properties of the extracted series of z-scores in the online case. Furthermore, for the online case additional investigations were made into the sensitivity of the methods to various parameters using hybrid synthetic data. This is important information that could be directly useful to practitioners when considering applying these methods to their own data sets. For offline detection the results of an experiment using fully synthetic data was also provided allowing properties of the data set to be adjusted directly providing deeper insight regarding how the method may generalise to other data sets. The next chapter will discuss the results of each of these experiments in detail.

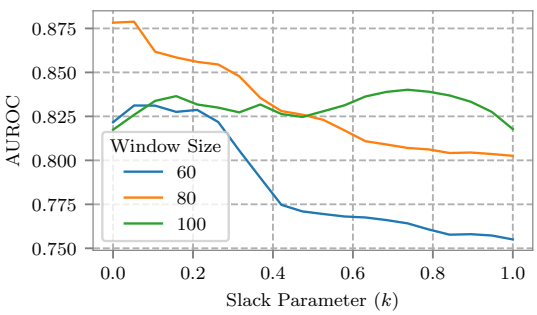


(A) No change time series

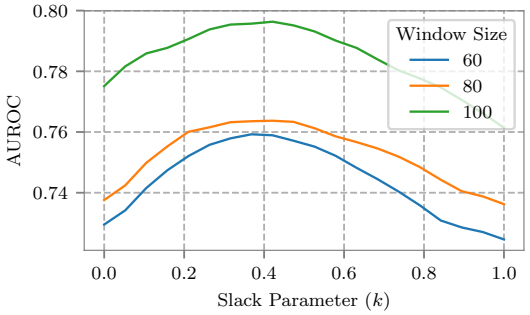


(B) Change time series

FIGURE 6.6: Histogram plot and autocorrelation plot for each of the z-score extraction methods investigated in this paper. (a) Joint (GMM Method), (b) Parametric, (c) Univariate Supervised. Evaluated on a no-change time series and a synthetic change time series from the Limpopo data set using MODIS Band 1.

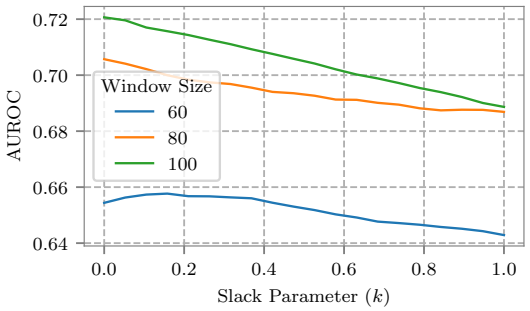


(A) GMM Joint, Limpopo

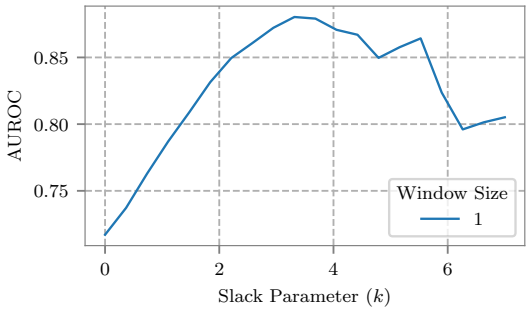


(B) GMM Joint, NSW

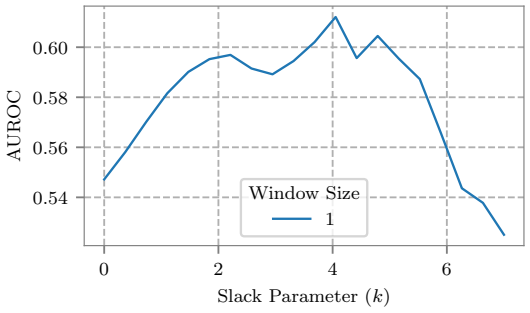
(C) Parametric, Limpopo



(D) Parametric, NSW

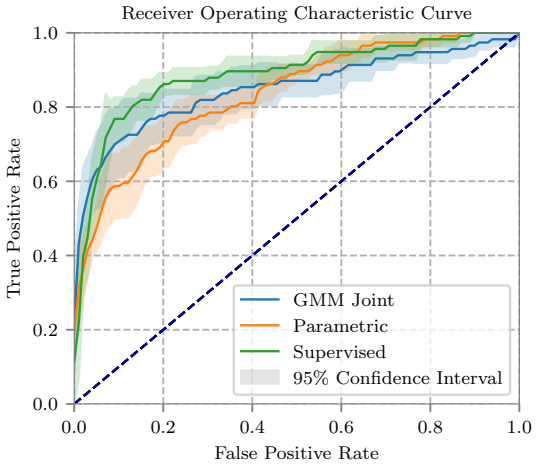


(E) Univariate Supervised, Limpopo

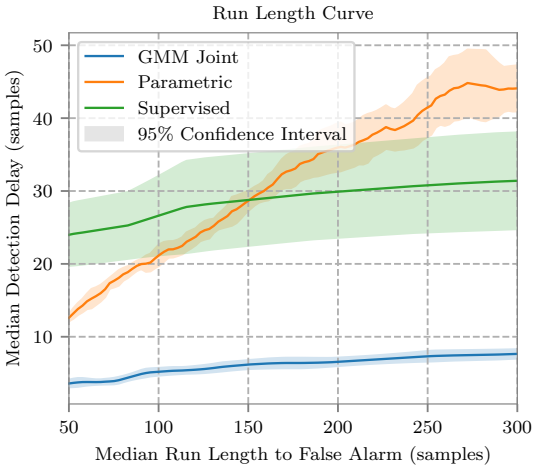


(F) Univariate Supervised, NSW

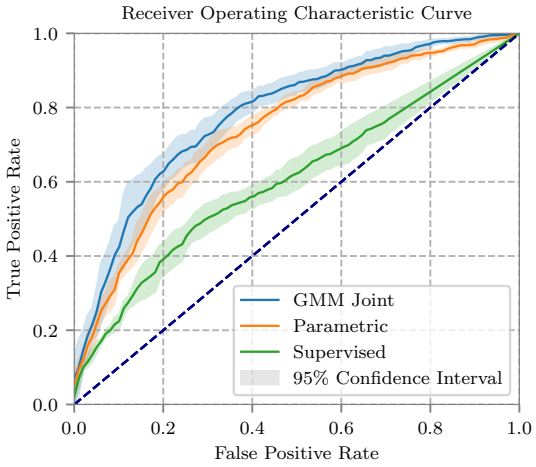
FIGURE 6.7: AUROC as a function of CUSUM slack parameter, k. Plotted for multiple selections of window size for MODIS Band 1.



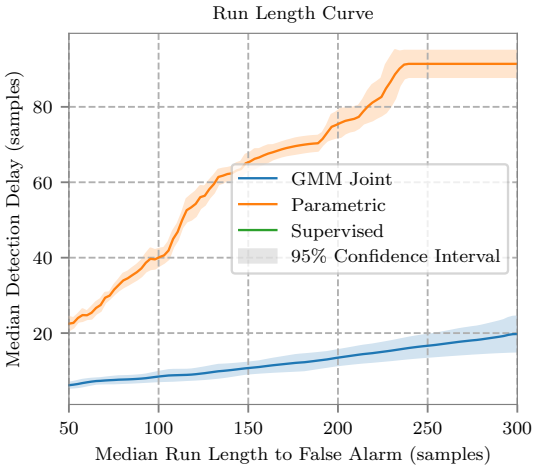
(A) ROC curves for the Limpopo data set band 1



(B) Detection Delay curves for the Limpopo data set band 1



(C) ROC curves for the NSW data set band 1



(D) Detection Delay curves for the NSW data set band 1

FIGURE 6.8: ROC and Detection Delay curves for each of the considered methods using MODIS Band 1. Curves were evaluated with a look-back window size of $W = 100$ and best slack parameter.

Chapter 7

Discussion

7.1 Introduction

The chapter begins with a thorough discussion of the results of the experiments for offline detection and change point estimation. This is followed by the online case. Each section includes a discussion of the observations from the experiments and these are considered in the context of how they support or refute the hypotheses of Chapters 4 and 5 and what this means for the overall aims of this study. Each section also includes a reflection on the limitations of the experiments conducted and suggestions for future research. A conclusion section is omitted from this chapter preferring instead to draw full conclusions in Chapter 8.

7.2 Offline Detection

Recall that the proposed hypothesis for offline change detection using a joint PDF was that the proposed method would perform as well as other methods despite requiring no additional parameters than the threshold. In order to test this the method of estimating per-class densities at each time step was evaluated in addition to the proposed three variations on the Mahalanobis distance detection approach. Also, for the Limpopo data set only, there were several existing sets of published change detection performance results with which to compare.

Comparing with existing results turned out to be somewhat problematic. This was due to the convention in the field of land cover change detection, and in many cases land cover classification as well, of publishing only the accuracy score with no estimate of confidence and often without accompanying TPR and FPR. This greatly limited the

number of studies that were suitable for comparison to only three methods from two publications. Even comparing these studies it is impossible to say with statistical significance if one method is superior to another. Despite this the results are compared in Table 6.3. Comparing with the Kalman filter approach of [6] it can at least be said that the Mahalanobis distance detection method using any of the covariance estimation methods exceed its performance on all of the evaluated metrics. To put this in the perspective of automation the Kalman filter used by this method requires the selection of a 1×3 initial state space vector, a 3×3 initial internal covariance matrix, a 3×3 process noise covariance matrix, the observation noise power and finally a decision threshold. Such is the complexity of selecting these parameters correctly that an entire paper [52] exists describing exactly that. It is shown that significant performance reduction can be expected if the parameters are not set correctly for a given problem. By employing the far less rigid model proposed in this work and leveraging the large swathes of existing data, one can likely expect improved performance with reduced overhead.

The non-linear detector based methods of [3] also listed in the comparison table exhibit exceptional performance on this data set especially in terms of FPR. This comes at the cost of an extremely complex algorithm. First the Kalman filter of [6] is applied to derive parameter time series (therefore requiring the additional parameters specified above), this is then used as input to a non-linear deterministic system which aims to amplify any permanent deviations in the derived time series. Finally the change/no-change decision is made by either defining an upper and lower threshold in the univariate case (two additional parameters) or training a Support Vector Machine (SVM) with a radial basis function kernel in the multivariate case (two additional parameters plus a training phase). While the given performance on this data set is superior, in practical terms this elaborate series of steps may not be worth the extra few percent of TPR for equivalent FPR when comparing the robust method and the single band non-linear detector method. The non-linear detector paper also explores the inclusion of additional bands and sees a further decrease in FPR. This is of particular interest and suggests that improved results may be obtained by including additional bands in the Mahalanobis distance based methods also. Another interesting experiment would be to use the derived z-score time series from the proposed online method as input to the non-linear detector phase as it ideally has similar properties to the parameter time series from the Kalman filter [3].

For practical purposes there is a strong argument for applying the method of Mahalanobis distance change detection due to the implementation simplicity and fewer number of parameters compared with the state-of-the-art methods applied to the Limpopo data set. There is insufficient data available from this comparison study alone to conclusively say that the proposed method is at least as good as other methods in the literature but the few results available suggest roughly equivalent performance with a significant decrease in complexity and number of parameters.

Considering now only the methods that were implemented as part of this study and their performance on both of the available data sets given in Tables 6.1 and 6.2. One thing that can be immediately noticed is the differences in difficulty between the two data sets. The performance is consistently poorer for the NSW data set in terms of AUROC in all cases except the uncorrelated method. Much more striking though is the comparison between different bands applied to these two data sets which can be seen in Appendix A Tables A.1 and A.2. For change detection of the conversion of vegetation to settlement in Limpopo the performance across all bands in terms of AUROC is quite consistent and, with the exception of EVI, the AUROC for the robust estimation method is above 0.99 irrespective of the band. In contrast for detecting forest thinning in New South Wales it is clear the NDVI significantly outperforms any of the other bands or indices. For this data set the choice of band has a far greater impact than the method for estimating the joint PDF. EVI is the next best but even this results in a significantly higher FPR and lower AUROC when compared with using the NDVI.

It is likely this relates to the nature of the change being considered. As shown in Fig. 6.1 the expansion of settlements in Limpopo involves the clearing of natural vegetation which is replaced by materials with very different spectral signatures such as gravel and corrugated steel roofing. This results in a change that is visible across all bands simultaneously. This is in contrast to the simulated change for New South Wales which was developed with the intention of controlling all variables with the exception of vegetation density. It is clear that the NDVI performs exactly as designed and results in an excellent proxy for leaf area index in this case. It is not known why the false positive rates are so high for the other bands considered including the component bands of the NDVI (bands 1 and 2) and the closely related EVI. This does however highlight the importance of band and index selection when targeting specific types of change. Any project attempting to

detect this kind of within-class change should consult existing studies or perform their own to identify the most sensitive bands and band combinations.

7.3 Offline change point estimation

With no published results for change point estimation with which to compare, the supervised uncorrelated method serves as the most suitable benchmark for the proposed methods. This method is well suited to comparison as it takes the alternative and more commonly seen approach of using land cover classes to reduce the variance of estimates of the land cover dynamics while ignoring within-signal correlation. For the Limpopo province this allowed estimating a separate uncorrelated model for the *vegetation* and *settlement* land cover classes. A similar approach was taken for the NSW data set but for the classes *high density vegetation* and *low density vegetation*. Additionally this model has the domain knowledge that changes are always from vegetation to settlement or high density to low density.

The error distribution plots of Fig. 6.4a and Fig. 6.4b present the results of the change point estimation experiments with the most clarity. Both the empirical estimator and the GMM estimator produce very tight error distributions on both data sets with 95% of errors being smaller than 15 samples. It is interesting to notice that the robust estimator which was highly effective for change detection performs comparatively poorly on both data sets for change point estimation. This observation led to further investigation in an experiment on fully synthetic data described below.

The uncorrelated supervised method did not perform well especially on the NSW data set. This is almost certainly due to the lack of clear definition between land cover classes in both data sets but especially the NSW set where the pairs of higher density and lower density pixels were selected relative to each other and hence have significant overlap at the class level. This highlights some of the potential problems with relying on land cover classes to reduce the variance in the model of the land cover dynamics.

The clear superiority of the empirical estimator compared with the robust estimator for change point estimation with the reverse being true for change detection, was a surprising result and this was investigated further using synthetic data. Fig. 6.5a and Fig. 6.5b were produced by evaluating change detection in terms of AUROC and change point

estimation in terms of MAE for various levels of between signal variation. Increasing the between signal variation results in an estimated covariance matrix with higher entropy. Recall that the robust estimator strives to eliminate outliers and produce a tighter covariance estimate by using only the subset of time series of size h that yields the minimum determinate of the covariance matrix. As shown by (6.10) this is also equal to the minimum entropy estimate.

Fig. 6.5a shows that the AUROC for a given SNR increases as the entropy decreases which is intuitively simple to understand. Reduced between-series variation means that step changes will produce a relatively greater change in the Mahalanobis distance leading to better separability between change and no-change series. Fig. 6.5b shows the inverse relationship where an increase in entropy results in a reduction (improvement) in the MAE for a given SNR. Some further intuition can be gained by looking at the plots of Mahalanobis distance as a function of change hypothesis for various levels of entropy as shown in Fig. 7.1.

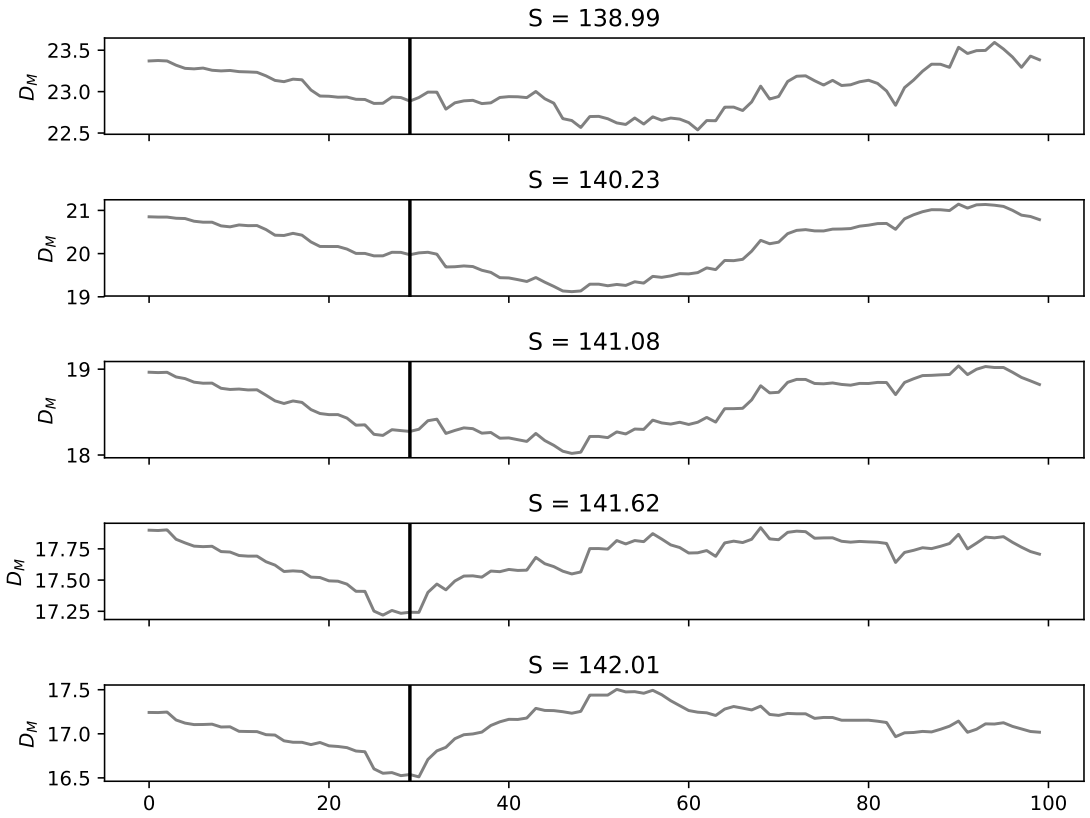


FIGURE 7.1: Plots of Mahalanobis distance, D_M as a function of change hypothesis for various levels of entropy for a simulated change time series. The real change point is given as a vertical line.

Several observations can be made from this series of plots. As the entropy increase the curves are shifted downward implying an overall smaller distance or higher likelihood for the signal under any of the change hypotheses. It also appears that the minima of the distance curve for more lower entropy models is biased toward the center of the series, the point where the covariance matrix has the fewest non-zero terms. This was verified by plotting the histogram of the position of the predicted change points for various entropy values as shown in Fig. 7.2. These histograms clearly show a bias toward predicting change points in the center that decreases as the entropy of the estimated joint PDF increases making it apparent that some kind of penalty term must be applied to the Mahalanobis distances to ensure the prediction distribution is always uniform. It is not immediately obvious what the optimal penalty should be and this is left as a further research problem. It is anticipated that a suitable derived penalty could allow the robust estimator to outperform the empirical estimator for change point estimation.

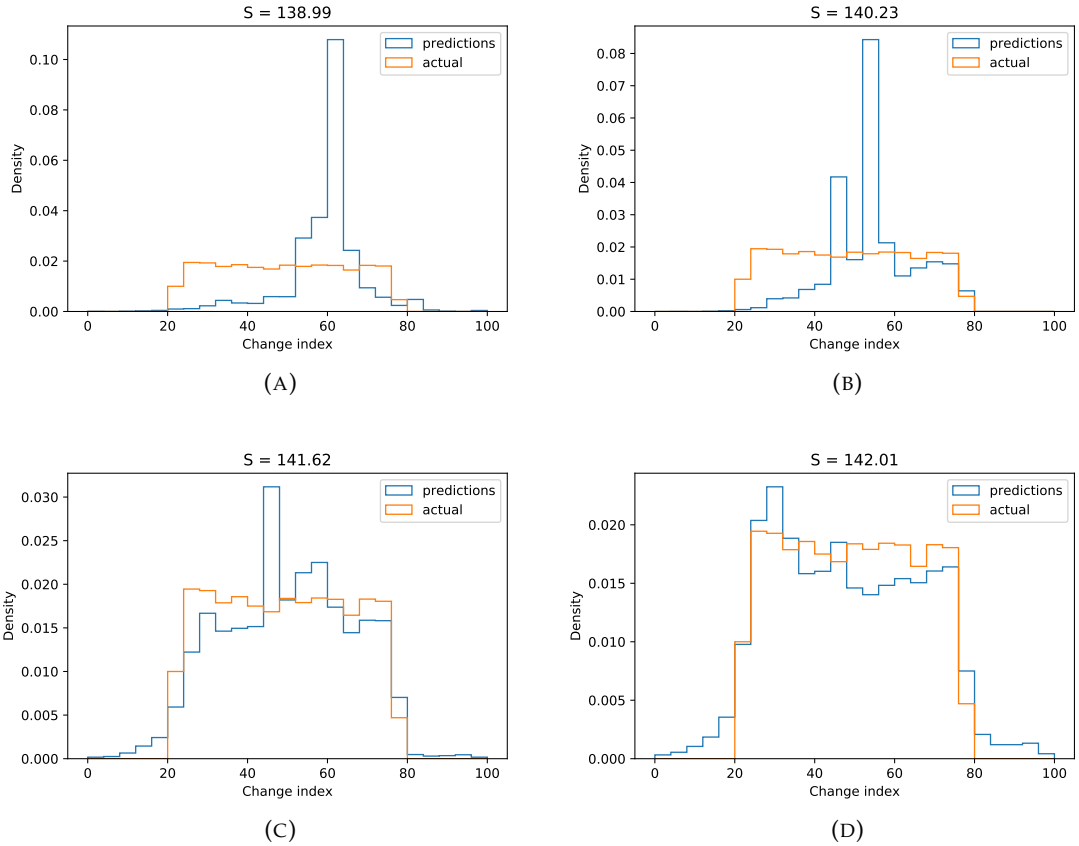


FIGURE 7.2: Histograms of actual and predicted change points on synthetic data for various levels of entropy.

7.4 Online Detection

The first set of experiments relates to testing the properties of the extracted z-score time series without considering change detection ability. Specifically testing the first part of the Chapter 5 hypothesis, that the series of z-scores is IID, unit Gaussian under no-change conditions. The results of applying the Kolmogorov-Smirnov test for zero-mean, unit-variance Gaussian distributed samples and the Ljung-Box test for independent samples are given for all bands in Tables A.7 and A.8. From the test statistic, D, and their p-values, p (K-S), it can be observed that the GMM joint method applied to bands 2, 5, 6 and 7 for the Limpopo data set and bands 1, 3, 5 and NDVI for the NSW data set support the hypothesis of Gaussian distributed z-score samples, as does band 2 using the parametric method. Additionally the parametric method applied to bands 2, 4 and NDVI for Limpopo and band 4 for NSW also appear significantly close to Gaussian distributed. Meeting this assumption indicates two things, that the predictions made by the method are unbiased and that the calibration of the predicted variance is close to correct. Both of these properties are required for the CUSUM detector to operate correctly. A predictor that produces biased estimates will have a monotonic curve for either C^+ or C^- if the slack parameter is smaller than the bias. This will of course result in frequent false alarms. Preventing this will require setting a larger slack parameter and thus reducing the sensitivity of the detector. This is supported by Fig. 6.7 which gives the AUROC performance as a function of slack parameter. It can be seen that a much higher slack ($k = 3$) is required by this method compared with the GMM and parametric methods ($k = 0.1$ and $k = 0.5$ respectively).

Well-calibrated variance estimates are required such that the confidence of the predictions is effectively captured. As made visible in Fig. 4.3 there are certain times of year when the natural variance is much higher and thus predictions should be made with less confidence/higher predicted variance. Failure to account for this will lead to artificially high z-scores during periods of naturally high variance also leading to an increase in false alarms. The other assumption required by the CUSUM method is that the input is uncorrelated. This was tested for using the Ljung-Box test for lags up to 10 samples with the results for all bands also included in Tables A.7 and A.8. From the test statistic,

Q, and their p-value, p (L-B), it is apparent the differences in correlation between the extracted z-score series using the GMM prediction method and the other two methods. For all bands with the exception of band 5 for the Limpopo data set and band 6 for the NSW data set, the test failed to reject the null hypothesis that the data is uncorrelated. Tests for each of the other methods for extracting z-score on any band yielded sufficient evidence to reject the null hypothesis and accept that the series is correlated. Correlated samples also have the potential to degrade the performance of the CUSUM detector which makes the assumption that the input data is IID under no-change conditions. It is not known what the measurable impact on change detection performance would be in the case of correlated data as it has the potential to lead to either overestimating or underestimating the values of the cumulative sums C_i^+ and C_i^- depending on the nature of the correlation.

Following on to the results of performance assessment, Tables 6.9 and 6.10 contain the AUROC and TPR at FPR=0.2 for the best performing bands by AUROC. On the Limpopo data set the supervised method performs extremely well almost equaling the AUROC of the offline methods. The ROC curve of Fig. 6.8a shows that for this data set the supervised method is consistently superior for any allowable FPR. This is not, however, reflected in the ability to detect change rapidly. To provide a single change metric the median DD was calculated for a RLFA of 200. This was chosen as it represents an acceptable expected time between false alarms of 1600 days, slightly over 4 years, at the MODIS MCD43A4 sample rate of 8 days. The median detection delay for the GMM joint method is less than half that of the other evaluated methods at 10 samples (80 days) compared with 22 samples (176 days) and 28 samples (224 days) for the parametric and supervised methods respectively. The comparison is more pronounced on the NSW data set. As discussed in section 7.2 the lack of well defined land cover classes make it very difficult for supervised methods to perform well on this problem. The reported results, especially the DD, for the supervised method suggest little better performance than triggering alarms randomly. In this case the GMM method applied to band 4 results in the best outcomes for each of the performance metrics. In particular the median detection delay is 13 samples (100 days) compared with 64 samples (512 days) for the parametric method. Recall that the detection delays are calculated relative to the start of a six month (23 sample) simulated blending between two different time series. Thus in both cases the GMM method is the only method able to detect the change before the transition is complete on average.

Both of these results lend support to the second aspect of the Chapter 5 hypothesis, that the series of z-scores derived from prediction using an estimated joint PDF is *biased when the time series experiences a change point. This can be used to rapidly detect the occurrence of change.* It is of course possible to obtain even shorter detection delays by reducing the threshold and also more frequent false alarms. The detection delay curves of Fig. 6.8 (or Appendix B for all bands) can be used to estimate the expected DD for a given RLFA. It can be seen that the curves are very flat for the GMM Joint method especially compared with the parametric method. Therefore there would be little gained in terms of DD by accepting a higher false alarm rate and in fact only a minor increase in expected DD would be experienced by increasing the threshold such that the expected RLFA was 300 samples (around 6.5 years). This was the highest RLFA that could be tested given the length of the available time series.

Comparing the results across all bands applied to the Limpopo data set included in Appendix A Table A.5 shows no particular band outperforming the others in terms of AUROC similar to the offline detection case. In terms of AUROC there is also no clear distinction as to which method performs best. This is not the case if the median DD at RLFA=200 is considered instead with the GMM Joint method producing the significantly smaller delays for every band. A similar trend can be seen for the NSW data set in terms of detection delay in Table A.6. What is different in this data set is the superiority of the GMM method for AUROC also, which as discussed prior, is likely due to the poor class definitions for this problem. Another notable observation is that the NDVI band which was particularly effective for offline detection on this data set does not appear to perform well for online detection. No explanation can be conceived only the inference that very different signal properties are required by the Mahalanobis based detection method and the online forecasting methods.

The influence of the parameters on the performance of online detection was of particular interest for this study given the focus on designing highly automated methods. The use of the CUSUM algorithm introduced an additional slack parameter as well as the size of the look-back window threshold and the threshold. To assess the influence of the slack parameter and the window size in a way that is invariant to the selection of threshold the AUROC was plotted for each method as a function of slack for various window sizes in Fig. 6.7 for Band 1 and Fig. B.9 for all bands.

The first observation that can be made is that there is very little to be gained by including a non-zero slack parameter for the GMM Joint or Parametric methods. In fact from looking at the collection of plots for all bands only band 7 and NDVI for the parametric method show an increase in AUROC of more than 0.1 for increasing the slack while the remainder are either marginally improved or AUROC performance decreases with increasing slack. This is not the case for the univariate supervised method which suffers a severe penalty for taking zero slack in every band across both data sets. In most cases reaching a peak AUROC at k between 2 and 4 with a steep decrease for values higher than the optimal setting. The difference between a zero and optimal slack is around 0.1 for each band in the Limpopo data set. Recall that the supervised method did not perform correctly on the NSW data set due to the lack of class separability and therefore the parameter curves for that problem are not of interest. The other point to note relates to the choice of window size. It is difficult to discern the exact relationship between window size and AUROC from these plots. In the majority of cases the larger look-back windows shows improved AUROC performance however in a number of band/method combinations this appears reversed. No conclusions regarding the influence of window size can be drawn from these experiments.

Although there are no published results applied to these exact data sets there are similar published results that can assist to place these results in the context of the literature. Online detection methods are applied to MODIS time series for detecting forest disturbance due to beetle infestation in [9]. This could be considered a similar problem to the NSW data set and uses similar method for simulating change time series. In this work the best reported median detection delay is 66 samples using a very similar parametric model based method. The work of [56] presents a different online method based on a Kalman filter to the same problem giving a median detection delay of 45 samples on simulated change time series. While these results relate to a different and likely more challenging problem the results in terms of detection delay are within a reasonable bound of the observed results on the conducted experiments.

Comparing with the results of [64] it is clear that the way forward for improving the detection delay for land cover change detection involves combining multiple, higher sample rate data sources. This recent study combines SAR data from Sentinel-1 and

ALOS PALSAR, with NDVI from Landsat-7 and Landsat-8. Using a supervised data fusion approach this method is able to detect total deforestation with a delay as short as 12 days with false alarm rate of around 10%. A comparable detection delay using MODIS would require detection in two samples which is infeasible with the SNR observed in practical problems.

7.5 Limitations of the study and future work

Reliance on existing literature for results with to compare with is, in theory, an excellent idea. It reduces the duplication between research groups and eliminates potential researcher bias. [1] In this field at least the convention of publishing performance metrics alone without quantifying the uncertainty makes meaningful comparisons impossible. This resulted in limiting the ability with which we could support or refute the hypotheses of Chapter 4 which related change detection and change point estimation performance to other current methods in the literature. It was important to include this as part of the hypothesis as without comparison it is impossible to state if progress has in fact been made toward the aim. In hindsight a better approach would be to comprehensively re-implement every recent method published. This is challenging because of the exact problems that this work aims to solve - the state of the art methods require numerous parameters and training stages making the studies difficult to recreate. Another alternative solution would be if the research community adopted a culture of sharing data and code similar to the open source software movement. This would promote higher quality research through enabling better comparisons, providing more data sets for evaluation and eliminate duplication of work.

Another factor that limits the generalisation of the results of chapter 4 is the reliance on hybrid synthetic change data. For the data set of New South Wales this was a necessity as no examples of real land cover change were observed. With additional resources of expertise and high resolution imagery it may have been possible to conduct a more detailed survey of this region to identify examples of real change but it is likely the number of examples would be too small to produce useful results. Furthermore this would only have been of use for experiments on change detection as it would still suffer the same

[1] It is documented that for algorithms with tunable parameters a subconscious bias can lead to more effort being directed to tuning the algorithms developed by the party conducting the experiment [86]

limitation of the Limpopo real change data regarding knowledge of the time of change. All of the experiments of change point estimation and online detection were conducted using hybrid synthetic change data due to limitations of identifying the time of change in real data. This also limits the ability to draw any conclusions regarding the expected performance of change point estimation or online detection to real change data. This is a limitation shared by every study of land cover change point estimation or online detection surveyed in the literature with individual studies either making use of synthetic data or omitting results regarding the change point estimation error altogether. A carefully created and curated data set with examples of real land cover change is something that would be of immense value to future researchers in this field.

A closely related limitation is the small numbers of regions on which the methods were evaluated. While care was taken to ensure the data sets were somewhat diverse containing instances of between-class and within-class changes, they are both located at similar inland, arid climates. From a passive remote sensing perspective this is very beneficial as these areas contain some of the lowest occurrences of cloud cover. At the same time this meant there was no need to consider techniques for dealing with significant periods of missing data. This is something that should be considered to ensure the methods could apply to truly diverse regions of land cover.

Finally, this study only considered individual MODIS bands in isolation or simple ratios such as NDVI. This was an intentional scope limitation as allowing multi-dimensional time series opens up a vast number of combinations and variations of the methods with which to compare. The question still remains if the performance of any of the proposed methods could be improved using a multi-dimensional time series approach. As its stands the method of estimating a joint PDF either over the entirety of the time series or in a finite look-back window can be trivially extended to the multi-band case. Several studies have shown improved performance by including additional bands in the model [3], [76], [87]. This serves as an excellent starting point for any future research.

Chapter 8

Conclusion

While the 'holy grail' of land cover change detection still remains an elusive goal this work provides some fresh ideas to the field and suggests some promising avenues for future research. In particular that the key to land cover change detection, both offline and online, is estimating an effective model of how a particular pixel time series should behave under no-change conditions. The parametric harmonic models popular in the literature were shown to be overly constraining and limited by only allowing samples within the candidate signal to be used for estimating parameters. It is shown that methods that make use of the correlations between samples in a time series estimated from a geographical region are a valid alternative.

For offline detection it was shown how a multivariate Gaussian model fit from an unlabeled training set of geographically similar pixels is an effective model. Change signals appear as outliers with respect to the Mahalanobis distance under the fitted PDF. Robust estimates of the covariance were found to yield improved change detection performance. We also present a method for applying the same model to change point estimation by testing multiple hypotheses about the correlation between samples before and after the change point. This method of change point estimation was shown to outperform a supervised learning method with additional domain knowledge of the land cover classes before and after a change. As an added bonus the proposed methods require no additional parameters except for the change decision threshold making them well suited for production systems. It was found that irrespective of the method, the selection of band has a significant impact on the change detection performance the optimal band choice depends on the properties of the land cover before and after the change.

For change point estimation it was found that the proposed method of partitioning

the estimated covariance matrix of the time series joint PDF and using multiple hypothesis testing was effective in identifying the time-of-change and could do so without requiring any parameters. This method was also applied to the multi-modal PDF estimated using a GMM by producing a single covariance matrix by partitioning and selecting the best fitting mixture component before and after the change point. It was also found that the method is sensitive to the entropy of the estimated covariance matrix with lower entropy models resulting in change estimates that are biased toward the center of the time series. The solution used in this study was to allow higher entropy PDFs by neglecting the robust covariance estimator in favor of the empirical or GMM estimator but it is hypothesised that a suitable penalty could be applied to allow lower entropy models to make unbiased predictions.

The method of using the joint PDF estimated over a fixed length look-back window to make forecasts of the next value of the time series was first evaluated by analyzing the resulting series errors between predicted and actual, normalised to produce z-scores. The hypothesis was tested of an ideal forecaster, which should produce a series of z-scores that is uncorrelated and unit Gaussian. It was found that using a GMM to estimate the joint and make predictions resulted in a series of z-scores that could be accepted as unit Gaussian for the majority of bands on each data set and could be accepted as uncorrelated for the vast majority. This results in a series of z-scores that has ideal properties under no-change conditions for applying classical change detection algorithms such as CUSUM.

Using the extracted z-score series for the online detection problem the joint estimation method was shown to exhibit significantly shorter median detection delays for a given median run length to false alarm when compared with the other considered methods. These results suggest the proposed method is a good candidate for land cover change detection system where rapid change detection is important. The idea of using prior samples as well as time series in the spatial neighbourhood to make forecasts for detecting change is powerful and there are many variations that could be devised under this framework to make use of multi-band or multi-sensor data.

These insights represent a step toward the aim stated at the beginning of this thesis and it is hoped that they lead toward the development of truly automated methods, by which the earths land cover can be effortlessly monitored in real time for dramatic changes. Data from more modern earth observation missions, such as VIIRS and

Sentinel-2, have has lower revisit time, higher resolution and higher SNR than ever before. It is up to the researchers to ensure that the algorithms and methods are constantly improving to make the best possible use of this incredible resource.

Bibliography

[1] B. P. Salmon, J. C. Olivier, W. Kleynhans, and K. J. Wessels, "Using the butterfly effect in a deterministic non-linear system to detect land cover change", in *Proc. IEEE Int. Geosci. Remote Sens. Symp*, Jul. 2014, pp. 4224–4227. DOI: 10.1109/IGARSS.2014.6947420.

[2] B. P. Salmon, J. C. Olivier, K. J. Wessels, W. Kleynhans, F. Van den Bergh, and K. C. Steenkamp, "Unsupervised land cover change detection: Meaningful sequential time series analysis", *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 4, no. 2, pp. 327–335, 2011.

[3] B. P. Salmon, D. Holloway, W. Kleynhans, J. C. Olivier, and K. J. Wessels, "Applying model parameters as a driving force to a deterministic nonlinear system to detect land cover change", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 12, pp. 7165–7176, 2017.

[4] T. L. Grobler, E. R. Ackermann, A. J. van Zyl, J. C. Olivier, W. Kleynhans, and B. P. Salmon, "Using Page's cumulative sum test on MODIS time series to detect land-cover changes", *Geoscience and Remote Sensing Letters, IEEE*, vol. 10, no. 2, pp. 332–336, 2013.

[5] W. Kleynhans, B. P. Salmon, J. C. Olivier, F. van den Bergh, K. J. Wessels, and T. Grobler, "Detecting land cover change using a sliding window temporal autocorrelation approach", in *Proc. IEEE Int. Geosci. Remote Sens. Symp*, Jul. 2012, pp. 6765–6768. DOI: 10.1109/IGARSS.2012.6352552.

[6] W. Kleynhans, J. C. Olivier, K. J. Wessels, B. P. Salmon, F. Van den Bergh, and K Steenkamp, "Detecting land cover change using an extended Kalman filter on MODIS NDVI time-series data", *Geoscience and Remote Sensing Letters, IEEE*, vol. 8, no. 3, pp. 507–511, 2011.

[7] W. Kleynhans, B. P. Salmon, J. C. Olivier, F. van den Bergh, K. J. Wessels, T. L. Grobler, and K. C. Steenkamp, "Land cover change detection using autocorrelation analysis on MODIS time-series data: Detection of new human settlements in the gauteng province of south africa", *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, no. 3, pp. 777–783, 2012.

[8] A. Anees, J. C. Olivier, M. O'Rielly, and J. Aryal, "Detecting beetle infestations in pine forests using MODIS NDVI time-series data", in *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*, IEEE, 2013, pp. 3329–3332.

[9] A. Anees and J. Aryal, "Near-real time detection of beetle infestation in pine forests using MODIS data", *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 7, no. 9, pp. 3713–3723, 2014.

[10] J. Rogan, J. Franklin, D. Stow, J. Miller, C. Woodcock, and D. Roberts, "Mapping land-cover modifications over large areas: A comparison of machine learning algorithms", *Remote Sensing of Environment*, vol. 112, no. 5, pp. 2272–2283, 2008.

[11] W. Olding, J. Olivier, B. Salmon, and W Kleynhans, "Unsupervised land cover change estimation using region covariance estimates", *IEEE Geoscience and Remote Sensing Letters*, 2018.

[12] W. C. Olding, J. C. Olivier, B. P. Salmon, and W. Kleynhans, "A forecasting approach to online change detection in land cover time series", *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 2019.

[13] W. J. Van Leeuwen, B. J. Orr, S. E. Marsh, and S. M. Herrmann, "Multi-sensor NDVI data continuity: Uncertainties and implications for vegetation monitoring applications", *Remote sensing of environment*, vol. 100, no. 1, pp. 67–81, 2006.

[14] R. E. Murphy, W. L. Barnes, A. I. Lyapustin, J. Privette, C. Welsch, F. DeLuccia, H. Swenson, C. F. Schueler, P. E. Ardanuy, and P. S. Kealy, "Using VIIRS to provide data continuity with MODIS", in *Geoscience and Remote Sensing Symposium, 2001. IGARSS'01. IEEE 2001 International*, IEEE, vol. 3, 2001, pp. 1212–1214.

[15] J. R. Jensen, *Remote sensing of the environment: An earth resource perspective 2/e*. Pearson Education India, 2009.

[16] W. G. Rees, *Physical principles of remote sensing*. Cambridge University Press, 2012.

[17] N. Landsat, *Science data users handbook*, 7.

[18] E. Vermote, N El Saleous, C. Justice, Y. Kaufman, J. Privette, L Remer, J. Roger, and D Tanre, "Atmospheric correction of visible to middle-infrared EOS-MODIS data over land surfaces: Background, operational algorithm and validation", *Journal of Geophysical Research: Atmospheres*, vol. 102, no. D14, pp. 17 131–17 141, 1997.

[19] A. H. Strahler, J. Muller, W Lucht, C Schaaf, T Tsang, F Gao, X Li, P. Lewis, and M. J. Barnsley, "MODIS BRDF/albedo product: Algorithm theoretical basis document version 5.0", *MODIS documentation*, vol. 23, no. 4, pp. 42–47, 1999.

[20] C. B. Schaaf, F. Gao, A. H. Strahler, W. Lucht, X. Li, T. Tsang, N. C. Strugnell, X. Zhang, Y. Jin, J.-P. Muller, *et al.*, "First operational BRDF, albedo nadir reflectance products from MODIS", *Remote Sens. Environ.*, vol. 83, no. 1-2, pp. 135–148, 2002.

[21] J. Rouse Jr, R. Haas, J. Schell, and D. Deering, "Monitoring vegetation systems in the great plains with ERTS", 1974.

[22] A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira, "Overview of the radiometric and biophysical performance of the MODIS vegetation indices", *Remote sensing of environment*, vol. 83, no. 1-2, pp. 195–213, 2002.

[23] U. of Concerned Scientists, *Ucs satellite database*, Database, Aug. 2017. [Online]. Available: http://www.ucsusa.org/satellite_database.

[24] C. Schaaf and Z W, "MCD43A4 MODIS/Terra+Aqua BRDF/albedo nadir BRDF adjusted L3 global - 500m v006.", in *NASA EOSDIS Land Processes DAAC*, 2015. DOI: https://doi.org/10.5067/modis/mcd43a4.006.

[25] C. Cao, J. Xiong, S. Blonski, Q. Liu, S. Uprety, X. Shao, Y. Bai, and F. Weng, "Suomi npp VIIRS sensor data record verification, validation, and long-term performance monitoring", *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 20, pp. 11–664, 2013.

[26] A. Singh, "Review article digital change detection techniques using remotely-sensed data", *International journal of remote sensing*, vol. 10, no. 6, pp. 989–1003, 1989.

[27] M Drusch, U Del Bello, S Carlier, O Colin, V Fernandez, F Gascon, B Hoersch, C Isola, P Laberinti, P Martimort, *et al.*, "Sentinel-2: ESA's optical high-resolution

mission for gmes operational services", *Remote Sensing of Environment*, vol. 120, pp. 25–36, 2012.

[28] R. F. Nelson, "Detecting forest canopy change due to insect activity using landsat mss", *Photogrammetric Engineering and Remote Sensing*, vol. 49, no. 9, pp. 1303–1314, 1983.

[29] P. J. Howarth and E. Boasson, "Landsat digital enhancements for change detection in urban environments", *Remote sensing of environment*, vol. 13, no. 2, pp. 149–160, 1983.

[30] P. J. Howarth and G. M. Wickware, "Procedures for change detection using landsat digital data", *International Journal of Remote Sensing*, vol. 2, no. 3, pp. 277–291, 1981.

[31] G. Byrne, P. Crapper, and K. Mayo, "Monitoring land-cover change by principal component analysis of multitemporal landsat data", *Remote sensing of Environment*, vol. 10, no. 3, pp. 175–184, 1980.

[32] W. A. Malila, "Change vector analysis: An approach for detecting forest changes with landsat", in *LARS symposia*, 1980, p. 385.

[33] T. Fung *et al.*, "An assessment of tm imagery for land-cover change detection.", *IEEE transactions on Geoscience and Remote Sensing*, vol. 28, no. 4, pp. 681–684, 1990.

[34] L Bruzzone and S. Serpico, "Detection of changes in remotely-sensed images by the selective use of multi-spectral information", *International Journal of Remote Sensing*, vol. 18, no. 18, pp. 3883–3888, 1997.

[35] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection", *IEEE Transactions on Geoscience and Remote sensing*, vol. 38, no. 3, pp. 1171–1182, 2000.

[36] J. Besag, "On the statistical analysis of dirty pictures", *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 259–302, 1986.

[37] R. S. Lunetta, J. F. Knight, J. Ediriwickrema, J. G. Lyon, and L. D. Worthy, "Land-cover change detection using multi-temporal MODIS NDVI data", *Remote Sens. Environ.*, vol. 105, no. 2, pp. 142–154, 2006.

[38] R. S. Lunetta, J. Ediriwickrema, D. M. Johnson, J. G. Lyon, and A. McKerrow, "Impacts of vegetation dynamics on the identification of land-cover change in a biologically complex community in north carolina, usa", *Remote Sensing of Environment*, vol. 82, no. 2-3, pp. 258–270, 2002.

[39] S. Boriah, V. Kumar, M. Steinbach, C. Potter, and S. Klooster, "Land cover change detection: A case study", in *Proc. Int. Conf. Knowl. Discovery and Data Mining*, ser. KDD '08, ACM, 2008, pp. 857–865, ISBN: 978-1-60558-193-4.

[40] R. E. Kennedy, W. B. Cohen, and T. A. Schroeder, "Trajectory-based change detection for automated characterization of forest disturbance dynamics", *Remote Sens. Environ.*, vol. 110, no. 3, pp. 370–386, 2007.

[41] R. E. Kennedy, Z. Yang, and W. B. Cohen, "Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms", *Remote Sens. Environ.*, vol. 114, no. 12, pp. 2897–2910, 2010.

[42] W. B. Cohen, Z. Yang, and R. Kennedy, "Detecting trends in forest disturbance and recovery using yearly landsat time series: 2. timesync—tools for calibration and validation", *Remote Sensing of Environment*, vol. 114, no. 12, pp. 2911–2924, 2010.

[43] H. Yin, D. Pflugmacher, R. E. Kennedy, D. Sulla-Menashe, and P. Hostert, "Mapping annual land use and land cover changes using MODIS time series", *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 7, no. 8, pp. 3421–3427, 2014.

[44] J. Verbesselt, A. Zeileis, and M. Herold, "Near real-time disturbance detection using satellite image time series", *Remote Sens. Environ.*, vol. 123, pp. 98–108, 2012.

[45] B. P. Salmon, J. C. Olivier, W. Kleynhans, K. J. Wessels, F. van den Bergh, and K. C. Steenkamp, "The use of a multilayer perceptron for detecting new human settlements from a time series of MODIS images", *Int. J. Appl. Earth Obs. Geoinf.*, vol. 13, no. 6, pp. 873–883, 2011.

[46] J Ronald Eastman, F. Sangermano, B. Ghimire, H. Zhu, H. Chen, N. Neeti, Y. Cai, E. A. Machado, and S. C. Crema, "Seasonal trend analysis of image time series", *Int. J. Remote Sens.*, vol. 30, no. 10, pp. 2721–2726, 2009.

[47] J. Verbesselt, R. Hyndman, G. Newnham, and D. Culvenor, "Detecting trend and seasonal changes in satellite image time series", *Remote Sens. Environ.*, vol. 114, no. 1, pp. 106 –115, 2010, ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2009.08.014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S003442570900265X.

[48] J. Verbesselt, R. Hyndman, A. Zeileis, and D. Culvenor, "Phenological change detection while accounting for abrupt and gradual trends in satellite image time series", *Remote Sens. Environ.*, vol. 114, no. 12, pp. 2970 –2980, 2010, ISSN: 0034-4257.

[49] A. Zeileis, C. Kleiber, W. Krämer, and K. Hornik, "Testing and dating of structural changes in practice", *Computational Statistics & Data Analysis*, vol. 44, no. 1-2, pp. 109–123, 2003.

[50] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory", *Journal of basic engineering*, vol. 83, no. 1, pp. 95–108, 1961.

[51] B. P. Salmon, W. Kleynhans, F. van den Bergh, J. C. Olivier, T. L. Grobler, and K. J. Wessels, "Land cover change detection using the internal covariance matrix of the extended Kalman filter over multiple spectral bands", *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 6, no. 3, pp. 1079–1085, Jul. 2013, ISSN: 1939-1404. DOI: 10.1109/JSTARS.2013.2241023.

[52] B. P. Salmon, W. Kleynhans, F. van den Bergh, J. C. Olivier, W. J. Marais, T. L. Grobler, and K. J. Wessels, "Meta-optimization of the extended Kalman filter's parameters through the use of the bias variance equilibrium point criterion", *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 8, pp. 5072–5087, 2014.

[53] W. Kleynhans, B. P. Salmon, K. J. Wessels, and J. C. Olivier, "Rapid detection of new and expanding human settlements in the Limpopo province of South Africa using a spatio-temporal change detection method", *International Journal of Applied Earth Observation and Geoinformation*, vol. 40, pp. 74–80, 2015.

[54] S. Aminikhanghahi and D. J. Cook, "A survey of methods for time series change point detection", *Knowledge and information systems*, vol. 51, no. 2, pp. 339–367, 2017.

[55] A. Anees and J. Aryal, "A statistical framework for near-real time detection of beetle infestation in pine forests using MODIS data", *Geoscience and Remote Sensing Letters, IEEE*, vol. 11, no. 10, pp. 1717–1721, 2014.

[56] A. Anees, J. Aryal, M. M. O'Reilly, and T. J. Gale, "A relative density ratio-based framework for detection of land cover changes in MODIS NDVI time series", *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 9, no. 8, pp. 3359–3371, 2016.

[57] E. S. Page, "Continuous inspection schemes", *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.

[58] D. C. Montgomery, *Introduction to statistical quality control*. John Wiley & Sons (New York), 2009.

[59] V. Chandola and R. R. Vatsavai, "A Gaussian process based online change detection algorithm for monitoring periodic time series", in *Proc. Int. Conf. Data Mining*, SIAM, 2011, pp. 95–106.

[60] B. Krishnamurthy, S. Sen, Y. Zhang, and Y. Chen, "Sketch-based change detection: Methods, evaluation, and applications", in *Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*, ACM, 2003, pp. 234–247.

[61] R. P. Adams and D. J. MacKay, "Bayesian online changepoint detection", *arXiv preprint arXiv:0710.3742*, 2007.

[62] V. Chandola and R. R. Vatsavai, "A scalable Gaussian process analysis algorithm for biomass monitoring", *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 4, no. 4, pp. 430–445, 2011.

[63] J. Reiche, S. de Bruin, D. Hoekman, J. Verbesselt, and M. Herold, "A bayesian approach to combine landsat and ALOS PALSAR time series for near real-time deforestation detection", *Remote Sensing*, vol. 7, no. 5, pp. 4973–4996, 2015.

[64] J. Reiche, E. Hamunyela, J. Verbesselt, D. Hoekman, and M. Herold, "Improving near-real time deforestation monitoring in tropical dry forests by combining dense sentinel-1 time series with landsat and ALOS-2 PALSAR-2", *Remote Sensing of Environment*, vol. 204, pp. 147–161, 2018.

[65] K. C. Steenkamp, K. J. Wessels, S. Archibald, and G. V. Maltitz, "Long-term phenology and variability of southern african vegetation", in *Proc. IEEE Int. Geosci. Remote Sens. Symp*, 2008, pp. 816–819. DOI: 10.1109/IGARSS.2008.4779474. [Online]. Available: https://doi.org/10.1109/IGARSS.2008.4779474.

[66] G. M. Foody, "Thematic map comparison", *Photogrammetric Engineering & Remote Sensing*, vol. 70, no. 5, pp. 627–633, 2004.

[67] I Masih, S Maskey, F. Mussá, and P Trambauer, "A review of droughts on the african continent: A geospatial and long-term perspective", *Hydrology and Earth System Sciences*, vol. 18, no. 9, pp. 3635–3649, 2014.

[68] P. J. Rousseeuw and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator", *Technometrics*, vol. 41, no. 3, pp. 212–223, 1999.

[69] P. J. Rousseeuw, "Multivariate estimation with high breakdown point", *Mathematical statistics and applications*, vol. 8, no. 283-297, p. 37, 1985.

[70] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other", *The annals of mathematical statistics*, pp. 50–60, 1947.

[71] R. S. Lunetta and C. D. Elvidge, *Remote sensing change detection*. Taylor & Francis, 1999.

[72] Z. G. Zhou and P. Tang, "Continuous anomaly detection in satellite image time series based on z-scores of season-trend model residuals", in *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*, IEEE, 2016, pp. 3410–3413.

[73] Y Ritov, "Decision theoretic optimality of the CUSUM procedure", *The Annals of Statistics*, pp. 1464–1469, 1990.

[74] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations", *J. Am. Stat. Assoc.*, vol. 53, no. 282, pp. 457–481, 1958.

[75] B. Salmon, W. Kleynhans, J. Olivier, and C. P. Schwegmann, "Improving features used for hyper-temporal land cover change detection by reducing the uncertainty in the feature extraction method", in *Proc. IEEE Int. Geosci. Remote Sens. Symp*, IEEE, 2017, pp. 1740–1743.

[76] W. Kleynhans *et al.*, "Detecting land-cover change using MODIS time-series data", PhD thesis, University of Pretoria, 2012.

[77] L. Stowe, E. McClain, R Carey, P Pellegrino, G. Gutman, P Davis, C Long, and S Hart, "Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data", *Advances in Space Research*, vol. 11, no. 3, pp. 51–54, 1991.

[78] P. Harrison, "Urbanization: The policies and politics of informal settlement in south africa: A historical perspective", *Africa Insight*, vol. 22, no. 1, pp. 14–22, 1992.

[79] B. P. Salmon *et al.*, "Improved hyper-temporal feature extraction methods for land cover change detection in satellite time series", PhD thesis, University of Pretoria, 2012.

[80] *Carbon farming initiative avoided deforestation 1.1 - methodology determination*, Australian Federal Register of Legislation, Mar. 2015.

[81] J. Benson, C. Allen, C Togher, and J Lemmon, "New south wales vegetation classification and assessment: Part 1 plant communities of the NSW western plains", *Cunninghamia*, vol. 9, no. 3, pp. 383–450, 2006.

[82] W. Olding, *Coordinates for pairs of neighbouring MODIS pixels containing dense and sparse land cover in new south wales, australia [data set]*, Zenodo, 2018. DOI: http://doi.org/10.5281/zenodo.1906658.

[83] A. Anees, J. Aryal, M. M. O'Reilly, T. J. Gale, and T. Wardlaw, "A robust multi-kernel change detection framework for detecting leaf beetle defoliation using landsat 7 etm+ data", *ISPRS J. Photogramm. Remote Sens.*, vol. 122, pp. 167–178, 2016.

[84] H. W. Lilliefors, "On the kolmogorov-smirnov test for normality with mean and variance unknown", *Journal of the American statistical Association*, vol. 62, no. 318, pp. 399–402, 1967.

[85] G. M. Ljung and G. E. Box, "On a measure of lack of fit in time series models", *Biometrika*, vol. 65, no. 2, pp. 297–303, 1978.

[86] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization", in *Advances in Neural Information Processing Systems*, 2011, pp. 2546–2554.

[87] J. Reiche, J. Verbesselt, D. Hoekman, and M. Herold, "Fusing landsat and SAR time series to detect deforestation in the tropics", *Remote Sensing of Environment*, vol. 156, pp. 276–293, 2015.

Appendix A

Experimental Data

This appendix contains the full tables of results obtained from each of the experiments. In each case a summarised table is included in-text with a reference to the full table in this section. This includes

- Complete set of results for offline detection applied to all MODIS bands on both data sets
- Complete set of MAE results for offline change estimation applied to all MODIS bands on both data sets
- Complete set of DD and AUROC results for online change estimation applied to all MODIS bands on both data sets
- Result of applying the Kolmogorov–Smirnov and Ljung-Box tests to the derived z-score time series for each MODIS band on both data sets

TABLE A.1: Test set results for offline change detection applied to real change data from the Limpopo settlement expansion set. Mean values are given in percentage with standard deviation across all cross validation folds given in parenthesis. Results are sorted descending by AUROC.

Band	Method	AUROC	TPR (%)	FPR (%)	Acc (%)
7	robust	0.999	93.91(12.18)	1.16(0.59)	98.59(0.37)
1	robust	0.998	98.26(3.48)	1.61(0.22)	98.38(0.17)
6	robust	0.998	97.39(5.21)	2.69(1.21)	97.32(0.97)
3	robust	0.997	99.13(1.74)	2.20(0.50)	97.87(0.48)
NDVI	robust	0.996	98.26(2.13)	4.08(1.86)	96.04(1.68)
5	robust	0.996	98.26(2.13)	3.23(2.71)	96.85(2.49)
2	robust	0.995	98.26(2.13)	4.12(2.57)	96.00(2.41)
4	robust	0.992	95.66(5.50)	4.84(1.15)	95.19(0.86)
5	empirical	0.989	93.92(6.51)	4.93(1.69)	95.02(1.36)
1	empirical	0.988	95.66(4.76)	5.73(1.86)	94.34(1.87)
7	empirical	0.987	93.99(8.49)	3.94(1.57)	95.95(1.59)
6	empirical	0.985	92.18(9.68)	8.07(2.72)	91.95(2.30)
2	empirical	0.984	94.79(10.43)	11.34(3.70)	88.97(3.07)
EVI	robust	0.980	99.13(1.74)	5.69(1.40)	94.55(1.37)
7	gmm	0.974	95.69(6.73)	9.72(2.86)	90.55(2.57)
3	empirical	0.972	86.14(10.82)	6.00(1.74)	93.61(1.55)
4	empirical	0.968	87.84(11.47)	7.53(3.41)	92.25(3.54)
1	gmm	0.965	93.09(5.24)	9.95(1.87)	90.20(1.95)
5	gmm	0.961	88.77(7.60)	6.63(0.95)	93.14(0.77)
6	gmm	0.955	87.04(13.77)	8.11(1.77)	91.65(1.92)
2	gmm	0.954	92.18(9.68)	16.89(1.96)	83.56(1.50)
NDVI	empirical	0.952	97.43(3.46)	20.03(6.79)	80.83(6.39)
3	gmm	0.926	79.19(21.57)	18.64(9.24)	81.26(7.92)
4	gmm	0.926	82.66(16.98)	14.56(2.74)	85.31(1.84)
NDVI	gmm	0.923	83.63(11.79)	21.64(9.54)	78.62(8.71)
EVI	empirical	0.919	92.18(9.28)	19.89(4.47)	80.71(4.37)
EVI	gmm	0.861	91.35(11.01)	38.53(7.25)	62.95(6.78)
NDVI	uncorrelated	0.726	98.26(2.13)	65.29(20.09)	37.86(18.98)
EVI	uncorrelated	0.703	93.92(6.51)	57.30(6.04)	45.23(5.86)
3	uncorrelated	0.605	92.17(8.43)	68.73(3.63)	34.28(3.31)
2	uncorrelated	0.567	95.69(4.76)	80.69(3.11)	23.08(2.86)
4	uncorrelated	0.535	83.86(16.33)	85.35(6.85)	18.06(7.02)
7	uncorrelated	0.518	91.40(9.87)	83.65(9.50)	20.06(8.66)
5	uncorrelated	0.518	92.21(6.41)	83.70(8.19)	20.06(7.50)
1	uncorrelated	0.511	87.23(10.79)	80.11(6.34)	23.21(6.48)
6	uncorrelated	0.504	90.64(8.47)	79.66(7.43)	23.81(7.27)

TABLE A.2: Test set results for offline change detection applied to the synthetic change data from the New South Wales devegetation set. Mean values are given in percentage with standard deviation across all cross validation folds given in parenthesis. Results are sorted descending by AUROC.

Band	Method	AUROC	TPR (%)	FPR (%)	Acc (%)
NDVI	robust	0.980	91.17(5.12)	1.40(0.86)	94.88(2.18)
NDVI	gmm	0.979	92.38(4.40)	2.31(0.75)	95.04(1.99)
NDVI	empirical	0.977	90.97(5.46)	1.91(0.97)	94.53(2.51)
EVI	robust	0.797	96.99(1.39)	46.34(3.04)	75.33(1.32)
EVI	gmm	0.772	97.09(2.51)	50.45(1.77)	73.32(0.93)
EVI	empirical	0.771	96.79(1.61)	48.45(2.98)	74.17(1.43)
NDVI	uncorrelated	0.757	95.09(5.04)	66.70(4.39)	64.19(1.13)
EVI	uncorrelated	0.699	81.24(32.41)	71.01(16.84)	55.12(8.02)
5	gmm	0.549	44.93(15.59)	41.62(15.93)	51.65(6.21)
7	gmm	0.549	88.57(6.13)	82.55(3.92)	53.01(3.21)
6	gmm	0.549	42.13(12.23)	33.60(2.92)	54.26(5.21)
5	empirical	0.547	62.49(30.61)	66.40(29.37)	48.04(2.64)
5	robust	0.545	51.55(6.88)	45.64(12.66)	52.96(6.19)
7	robust	0.543	91.17(6.96)	86.26(4.22)	52.46(2.62)
2	empirical	0.537	77.63(26.03)	80.84(21.94)	48.40(2.22)
6	robust	0.537	50.05(29.91)	50.45(29.29)	49.80(5.82)
7	empirical	0.535	86.26(12.22)	85.76(8.61)	50.25(3.52)
6	empirical	0.535	57.97(31.05)	61.58(30.41)	48.19(5.31)
1	robust	0.535	66.70(29.97)	73.02(22.77)	46.84(4.99)
2	robust	0.534	73.12(20.02)	74.32(19.59)	49.40(2.99)
3	robust	0.533	38.21(32.20)	39.42(31.68)	49.40(3.05)
1	empirical	0.532	72.52(24.84)	78.13(21.68)	47.19(3.85)
1	gmm	0.532	51.76(33.41)	54.16(32.40)	48.80(2.61)
4	robust	0.532	84.95(29.28)	88.97(17.28)	47.99(6.02)
3	empirical	0.532	27.28(25.85)	28.59(24.63)	49.35(2.57)
4	gmm	0.529	42.63(45.58)	42.43(46.44)	50.10(0.68)
4	empirical	0.526	46.34(10.05)	41.22(21.82)	52.56(6.02)
3	gmm	0.526	22.37(9.15)	16.85(4.55)	52.76(2.81)
2	uncorrelated	0.525	85.86(11.81)	85.26(8.72)	50.30(3.54)
3	uncorrelated	0.525	85.56(9.66)	86.06(5.51)	49.75(6.25)
2	gmm	0.524	79.44(24.79)	82.65(22.45)	48.40(1.68)
1	uncorrelated	0.516	74.32(12.05)	82.95(7.55)	45.69(3.35)
5	uncorrelated	0.512	88.37(6.85)	87.16(2.29)	50.60(2.62)
6	uncorrelated	0.509	70.51(8.03)	71.41(5.08)	49.55(3.87)
7	uncorrelated	0.507	70.01(7.83)	75.43(8.09)	47.29(3.71)
4	uncorrelated	0.504	76.13(15.53)	82.35(9.09)	46.89(5.52)

TABLE A.3: Comparison of mean error (Mean) and Mean Absolute Error (MAE) between estimated and known change point on synthetic change data from the Limpopo data set. Standard deviation across all cross validation folds given in parenthesis. Errors are given as the numer of samples.

Band	Method	Mean	MAE
1	gmm	-0.26	5.38(7.57)
6	gmm	-0.54	5.74(7.05)
7	gmm	-0.24	5.76(5.90)
5	gmm	-0.06	5.80(7.93)
2	gmm	-1.62	6.16(8.50)
4	gmm	-0.69	6.23(12.53)
5	empirical	-0.26	6.92(10.82)
3	gmm	1.96	7.00(11.67)
6	empirical	-1.26	7.24(9.31)
7	empirical	0.46	7.36(13.92)
2	empirical	-1.11	7.39(9.33)
3	empirical	3.16	8.34(17.52)
4	empirical	0.93	8.61(19.07)
1	empirical	0.64	8.80(22.12)
EVI	gmm	1.49	10.71(27.77)
NDVI	empirical	0.46	13.24(29.62)
7	robust	-1.40	13.52(27.47)
1	robust	2.45	15.49(33.79)
EVI	empirical	1.05	15.69(33.26)
5	robust	-5.15	16.53(29.56)
NDVI	gmm	1.52	16.72(28.38)
4	robust	1.46	17.90(40.73)
2	robust	-4.32	19.48(36.82)
6	robust	-3.10	20.02(37.99)
3	robust	3.47	22.07(43.05)
EVI	robust	5.24	26.06(42.05)
NDVI	robust	-3.44	28.56(44.49)
1	uncorrelated	6.86	52.44(88.97)
2	uncorrelated	4.81	56.89(88.24)
4	uncorrelated	-5.95	58.49(90.35)
3	uncorrelated	-14.31	72.05(96.03)
NDVI	uncorrelated	14.59	72.39(93.85)
7	uncorrelated	-14.86	76.42(99.39)
EVI	uncorrelated	-59.85	85.45(84.46)
5	uncorrelated	-12.93	87.83(108.85)
6	uncorrelated	-21.71	101.09(112.66)

TABLE A.4: Comparison of mean error (Mean) and Mean Absolute Error (MAE) between estimated and known change point on synthetic change data from the NSW data set. Standard deviation across all cross validation folds given in parenthesis. Errors are given as the numer of samples.

Band	Method	Mean	MAE
NDVI	gmm	0.62	5.52(6.82)
NDVI	empirical	0.62	6.00(6.98)
5	empirical	-1.75	6.75(8.49)
1	empirical	-1.23	7.43(14.37)
5	gmm	-0.98	7.58(15.55)
4	gmm	0.17	7.73(12.27)
3	gmm	-1.06	8.10(9.89)
7	empirical	0.17	8.39(18.75)
4	empirical	-2.01	9.25(17.37)
6	empirical	-1.98	9.36(16.45)
7	gmm	-0.95	9.57(17.64)
3	empirical	0.55	9.61(14.31)
2	empirical	-2.47	10.15(19.39)
2	gmm	-2.93	10.63(17.17)
6	gmm	-4.27	12.99(25.44)
1	gmm	-1.97	15.35(30.18)
7	robust	-3.67	24.87(43.90)
2	robust	4.88	25.30(40.38)
3	robust	3.21	26.45(44.20)
6	robust	1.48	26.60(46.56)
1	robust	-0.42	26.70(50.24)
5	robust	8.30	27.34(44.01)
4	robust	9.18	27.62(45.94)
EVI	gmm	7.83	36.87(60.14)
EVI	empirical	3.53	38.21(51.57)
NDVI	robust	0.89	48.09(58.50)
EVI	robust	19.31	71.21(82.23)
NDVI	uncorrelated	-1.03	110.11(122.57)
EVI	uncorrelated	-132.84	151.72(102.71)
4	uncorrelated	200.36	200.36(108.79)
3	uncorrelated	200.36	200.36(108.79)
2	uncorrelated	200.36	200.36(108.79)
1	uncorrelated	200.36	200.36(108.79)
7	uncorrelated	-215.64	215.64(108.79)
6	uncorrelated	-222.64	222.64(108.79)
5	uncorrelated	-222.64	222.64(108.79)

TABLE A.5: Results for detection of settlement expansion in Limpopo Province. '-' designates cases where the estimated median DD was longer than the time series and a therefore a result could not be obtained.

Method	AUROC	Band	TPR at FPR=0.2	DD at RLFA=200	Threshold at RLFA=200
Supervised	0.90	1	86.20(3.29)	28.00	5.52
GMM Joint	0.87	7	75.94(6.65)	10.00	23.58
Parametric	0.86	4	87.03(3.97)	22.00	68.73
Supervised	0.82	4	71.63(10.92)	14.00	4.01
Supervised	0.82	3	76.81(8.15)	25.00	5.52
Supervised	0.82	7	71.49(8.23)	52.00	7.02
Parametric	0.81	5	68.88(6.67)	36.00	79.26
Supervised	0.81	2	65.58(6.87)	171.00	7.02
GMM Joint	0.79	6	75.91(5.65)	8.00	19.06
Parametric	0.79	1	70.80(7.84)	43.00	92.81
GMM Joint	0.77	NDVI	56.01(20.31)	17.00	21.00
Supervised	0.77	NDVI	61.16(9.67)	-	7.00
GMM Joint	0.77	1	77.68(8.02)	7.00	20.57
Parametric	0.76	2	63.80(6.97)	33.00	77.76
GMM Joint	0.75	4	69.06(8.39)	7.00	20.57
Parametric	0.72	3	70.72(7.30)	32.00	64.21
GMM Joint	0.71	3	67.32(8.98)	9.00	20.57
Supervised	0.69	6	51.81(13.44)	106.00	2.51
GMM Joint	0.68	5	61.20(8.72)	10.00	20.57
Parametric	0.66	6	52.57(8.29)	42.00	79.26
Supervised	0.65	5	53.55(8.10)	274.00	2.51
Parametric	0.61	7	53.44(8.43)	62.00	101.84
Parametric	0.59	NDVI	44.06(14.60)	72.00	82.00
GMM Joint	0.57	2	52.61(12.17)	11.00	20.57

TABLE A.6: Results for detection of vegetation thinning in New South Wales. '-' designates cases where the estimated median DD was longer than the time series and a therefore a result could not be obtained.

Method	AUROC	Band	TPR at FPR=0.2	DD at RLFA=200	Threshold at RLFA=200
GMM Joint	0.84	4	75.64(2.64)	13.00	18.07
GMM Joint	0.81	5	68.11(2.69)	13.00	17.77
GMM Joint	0.81	6	64.50(6.20)	15.00	19.14
GMM Joint	0.79	3	53.88(11.36)	24.00	19.66
Parametric	0.76	6	54.76(5.24)	64.00	127.99
Parametric	0.76	1	56.17(4.25)	68.00	140.79
Parametric	0.76	5	52.05(4.78)	69.00	119.20
GMM Joint	0.74	1	62.99(5.13)	16.00	19.14
Parametric	0.74	7	51.65(3.98)	66.00	136.52
GMM Joint	0.73	2	29.21(11.33)	10.00	8.47
Parametric	0.70	4	64.99(4.68)	64.00	93.84
Parametric	0.70	2	43.63(4.49)	84.00	130.12
Parametric	0.70	NDVI	46.35(5.93)	84.00	115.74
Supervised	0.68	2	46.34(2.69)	315.00	6.34
Supervised	0.67	5	40.53(4.03)	-	6.59
Supervised	0.65	6	38.92(6.38)	-	7.40
GMM Joint	0.65	7	48.52(14.22)	23.00	20.21
Parametric	0.65	3	51.45(2.05)	84.00	84.03
Supervised	0.64	1	39.31(4.89)	-	9.54
Supervised	0.61	7	35.50(5.29)	-	11.67
Supervised	0.61	NDVI	31.30(5.75)	-	5.66
Supervised	0.59	4	36.91(4.08)	-	6.34
Supervised	0.56	3	29.79(2.83)	-	4.73
GMM Joint	0.53	NDVI	32.29(7.09)	-	101.75

TABLE A.7: Table of test statistics and p-values for the Kolmogorov–Smirnov and Ljung-Box applied to the derived z-scores for the Limpopo data set. Values in boldface indicate cases where the null could not be rejected.

method	band	D	p (K-S)	Q	p (L-B)
GMM Joint	1	0.075	0.008	9.5	**0.48**
GMM Joint	2	0.037	**0.5**	1e+01	**0.41**
GMM Joint	NDVI	0.07	0.016	1.1e+01	**0.34**
GMM Joint	7	0.038	**0.47**	1.7e+01	**0.071**
GMM Joint	6	0.033	**0.66**	8.2	**0.61**
GMM Joint	5	0.031	**0.74**	2.1e+01	0.02
GMM Joint	4	0.061	0.05	1.4e+01	**0.16**
GMM Joint	3	0.07	0.016	1.2e+01	**0.3**
Parametric	1	0.099	0.00012	7.3e+02	8.8e-150
Parametric	2	0.058	**0.066**	3.8e+02	3e-76
Parametric	NDVI	0.061	**0.052**	1.1e+03	3.4e-233
Parametric	7	0.11	5.7e-06	1.1e+03	2.3e-219
Parametric	6	0.093	0.00036	7.2e+02	7e-148
Parametric	5	0.066	0.027	3.5e+02	6.8e-69
Parametric	4	0.052	**0.13**	2.2e+02	5.6e-42
Parametric	3	0.075	0.0077	1.7e+02	4.9e-31
Unsupervised	3	0.33	0.0	2.1e+02	1.9e-40
Unsupervised	1	0.25	0.0	5.4e+02	5.3e-109
Unsupervised	5	0.38	0.0	1.8e+02	4.2e-33
Unsupervised	6	0.25	0.0	4e+02	5.5e-79
Unsupervised	7	0.3	0.0	6.1e+02	4.4e-124
Unsupervised	NDVI	0.28	0.0	1.1e+03	1.1e-234
Unsupervised	2	0.36	0.0	1.6e+02	3.4e-29
Unsupervised	4	0.38	0.0	1.5e+02	6.4e-27

TABLE A.8: Table of test statistics and p-values for the Kolmogorov–Smirnov and Ljung-Box applied to the derived z-scores for the NSW data set. Values in boldface indicate cases where the null could not be rejected.

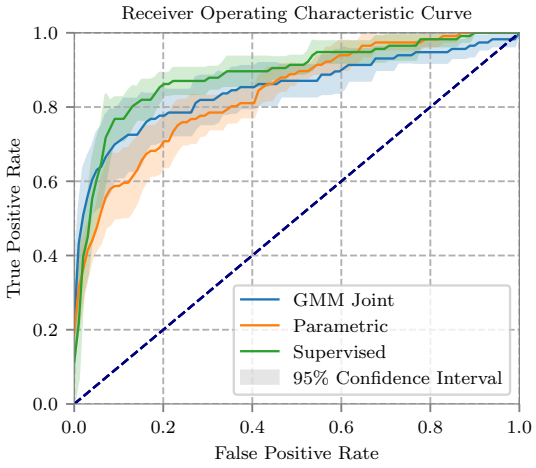
method	band	D	p (K-S)	Q	p (L-B)
GMM Joint	1	0.032	**0.89**	3.7	**0.96**
GMM Joint	2	0.091	0.0089	1.2e+01	**0.28**
GMM Joint	NDVI	0.05	**0.38**	7.3	**0.7**
GMM Joint	7	0.11	0.0012	1.5e+01	**0.13**
GMM Joint	6	0.076	0.046	2.8e+01	0.0019
GMM Joint	5	0.066	**0.11**	6.3	**0.79**
GMM Joint	4	0.09	0.0094	3.6	**0.96**
GMM Joint	3	0.062	**0.16**	7.8	**0.65**
Parametric	1	0.18	1.5e-09	1.7e+03	0.0
Parametric	2	0.22	0.0	1.5e+03	0.0
Parametric	NDVI	0.12	0.0003	1.2e+03	1.3e-253
Parametric	7	0.18	1.1e-09	1.8e+03	0.0
Parametric	6	0.19	2e-10	1.6e+03	0.0
Parametric	5	0.24	0.0	1.4e+03	3.8e-302
Parametric	4	0.061	**0.17**	3.8e+02	3.2e-75
Parametric	3	0.094	0.0064	3.3e+02	4.6e-65
Unsupervised	3	0.55	0.0	2.8e+02	1.6e-54
Unsupervised	1	0.81	0.0	9.6e+02	4.2e-200
Unsupervised	5	0.62	0.0	1.6e+03	0.0
Unsupervised	6	0.71	0.0	1.5e+03	0.0
Unsupervised	7	0.76	0.0	1.6e+03	0.0
Unsupervised	NDVI	0.74	0.0	1e+03	2.8e-212
Unsupervised	2	0.59	0.0	1.6e+03	0.0
Unsupervised	4	0.65	0.0	1.5e+02	1.4e-27

Appendix B

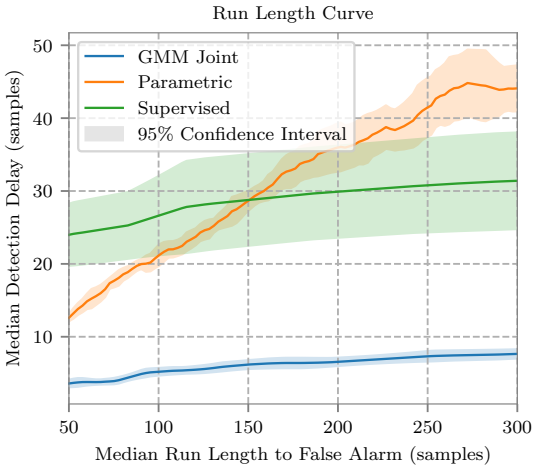
Supplementary Plots

This appendix contains the plots of results that could not be included in-text. This includes the following

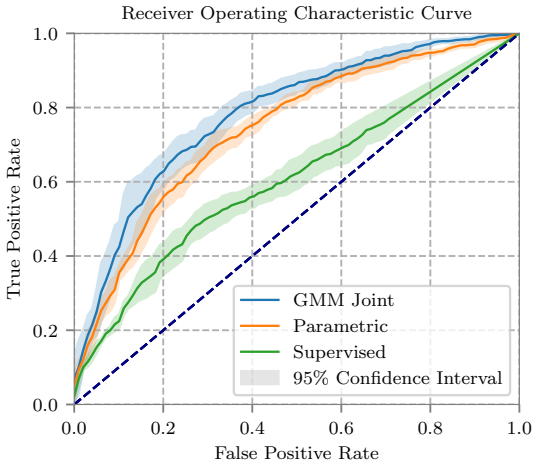
- ROC curves online change detection using each available MODIS band.
- Detection delay curves for online detection using each available MODIS band.
- Parameter sensitivity plots for AUROC as a function of slack parameter, k, for each method applied to each MODIS band.



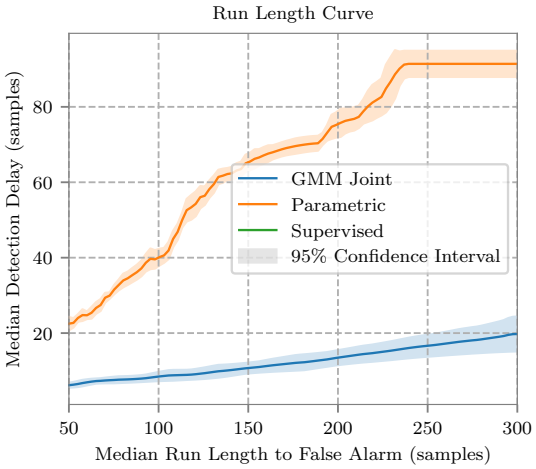
(A) ROC curves for the Limpopo data set band 1



(B) Detection Delay curves for the Limpopo data set band 1

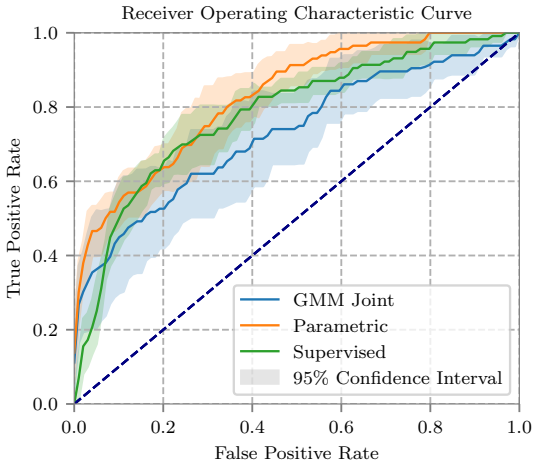


(C) ROC curves for the NSW data set band 1

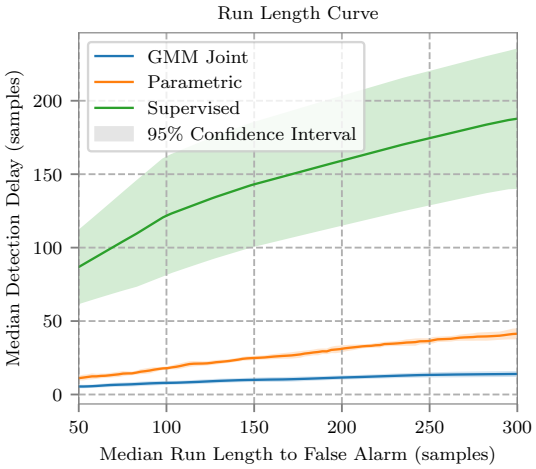


(D) Detection Delay curves for the NSW data set band 1

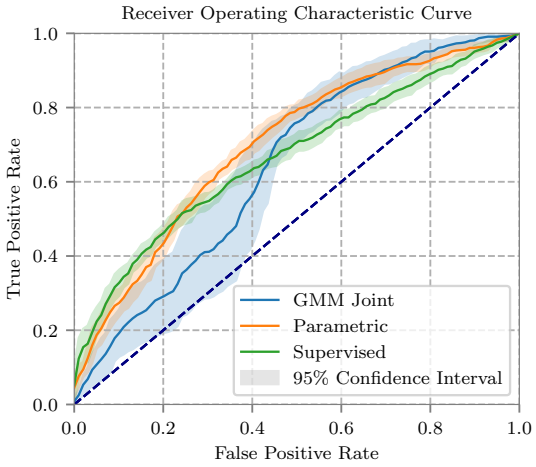
FIGURE B.1: ROC and Detection Delay curves for each of the considered methods using MODIS Band 1. Curves were evaluated with a look-back window size of $W = 100$ and best slack parameter.



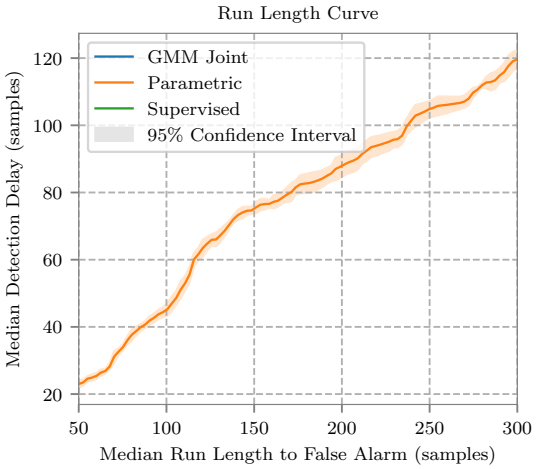
(A) ROC curves for the Limpopo data set band 2



(B) Detection Delay curves for the Limpopo data set band 2

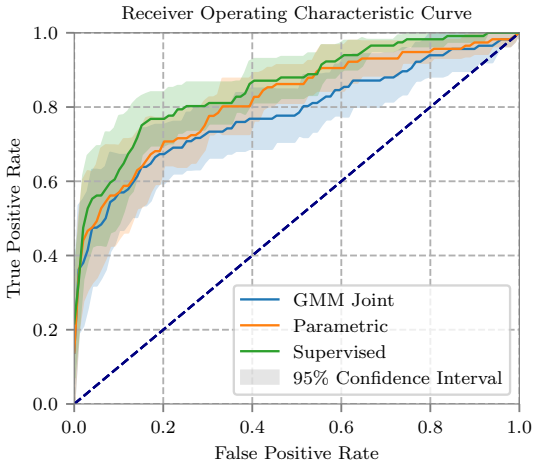


(C) ROC curves for the NSW data set band 2

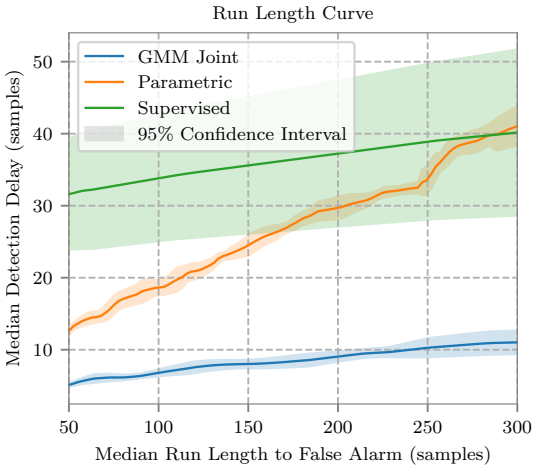


(D) Detection Delay curves for the NSW data set band 2

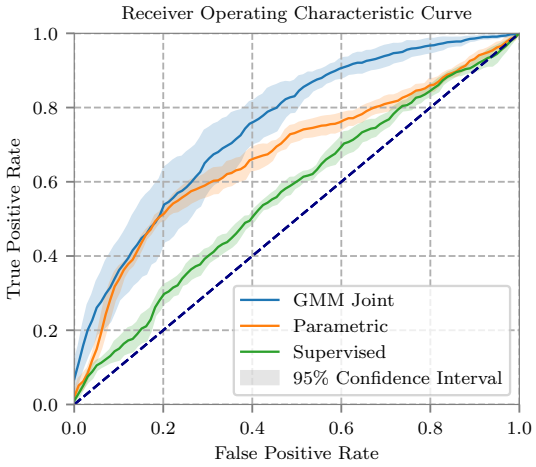
FIGURE B.2: ROC and Detection Delay curves for each of the considered methods using MODIS Band 2. Curves were evaluated with a look-back window size of $W = 100$ and best slack parameter.



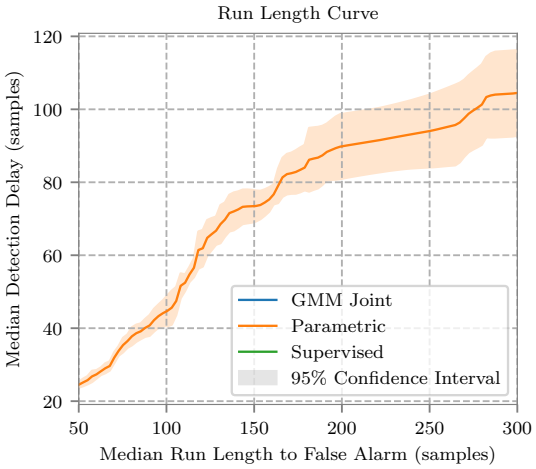
(A) ROC curves for the Limpopo data set band 3



(B) Detection Delay curves for the Limpopo data set band 3

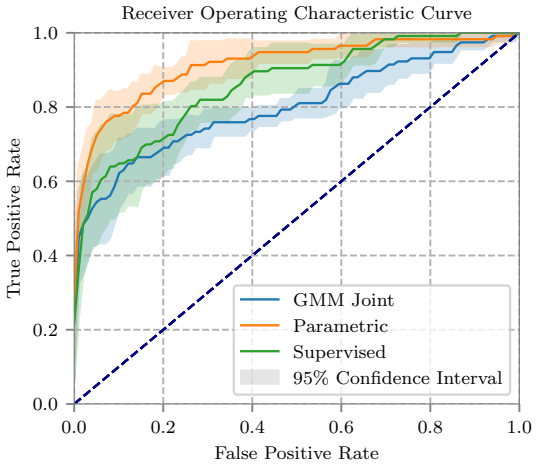


(C) ROC curves for the NSW data set band 3

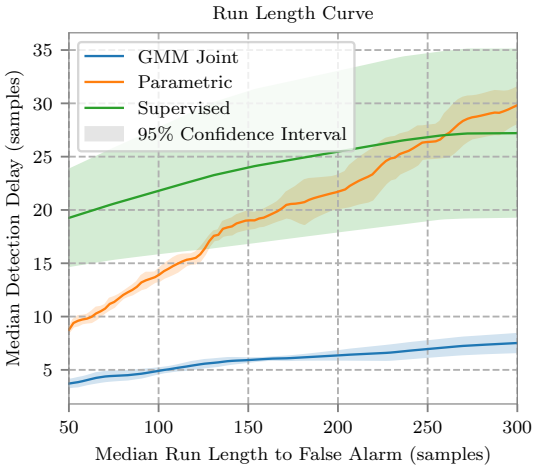


(D) Detection Delay curves for the NSW data set band 3

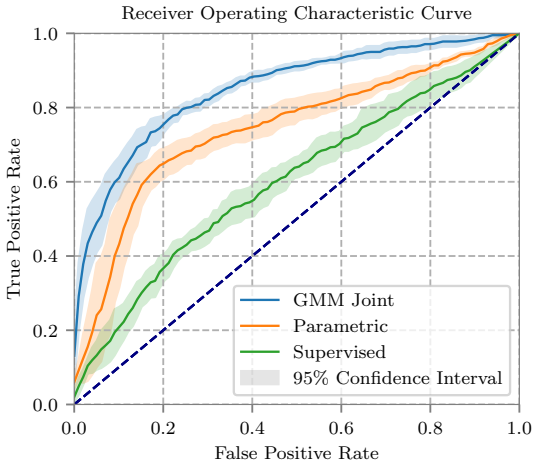
FIGURE B.3: ROC and Detection Delay curves for each of the considered methods using MODIS Band 3. Curves were evaluated with a look-back window size of $W = 100$ and best slack parameter.



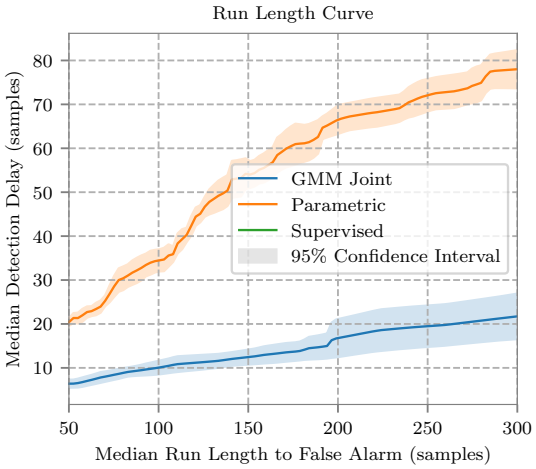
(A) ROC curves for the Limpopo data set band 4



(B) Detection Delay curves for the Limpopo data set band 4

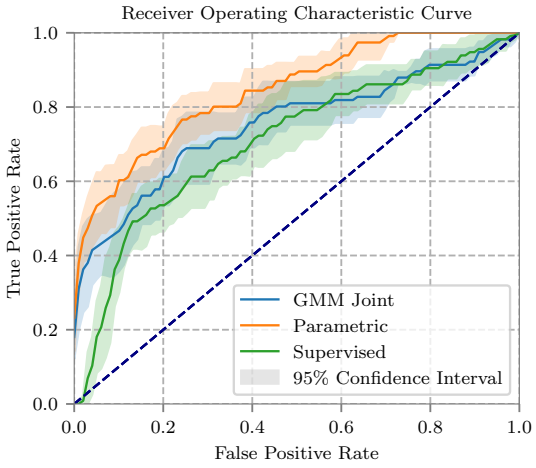


(C) ROC curves for the NSW data set band 4

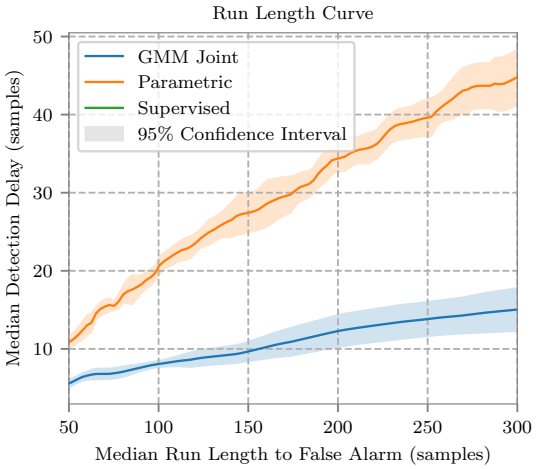


(D) Detection Delay curves for the NSW data set band 4

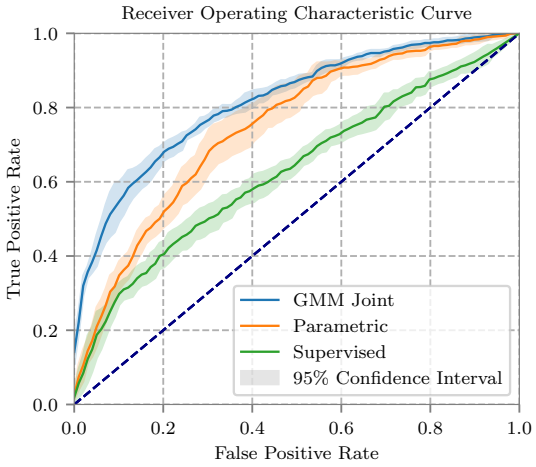
FIGURE B.4: ROC and Detection Delay curves for each of the considered methods using MODIS Band 4. Curves were evaluated with a look-back window size of $W = 100$ and best slack parameter.



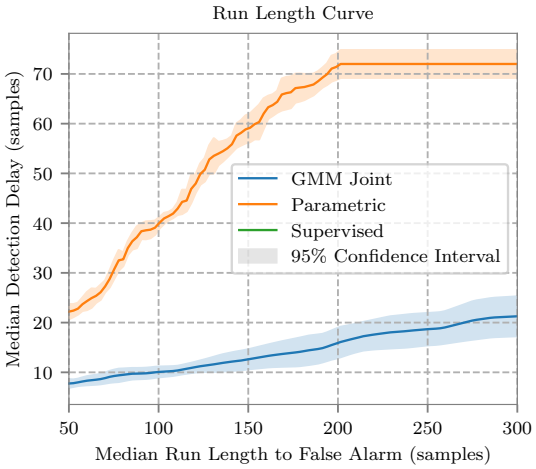
(A) ROC curves for the Limpopo data set band 5



(B) Detection Delay curves for the Limpopo data set band 5

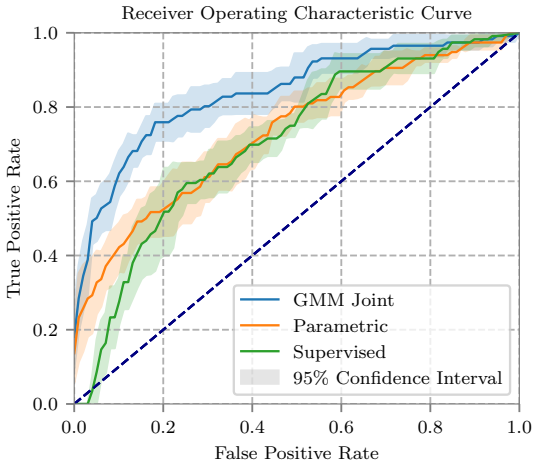


(C) ROC curves for the NSW data set band 5

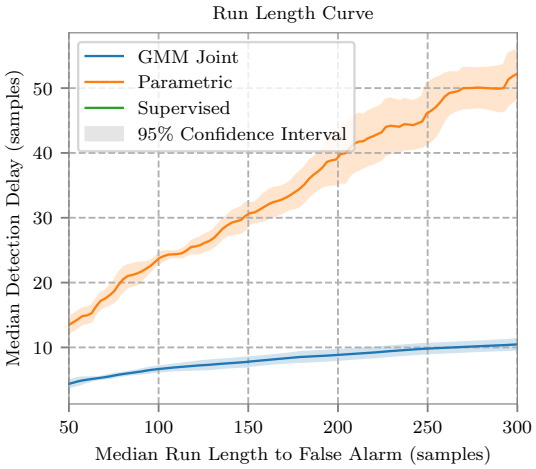


(D) Detection Delay curves for the NSW data set band 5

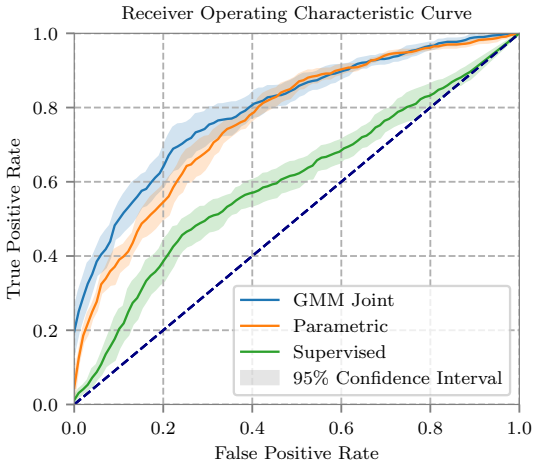
FIGURE B.5: ROC and Detection Delay curves for each of the considered methods using MODIS Band 5. Curves were evaluated with a look-back window size of $W = 100$ and best slack parameter.



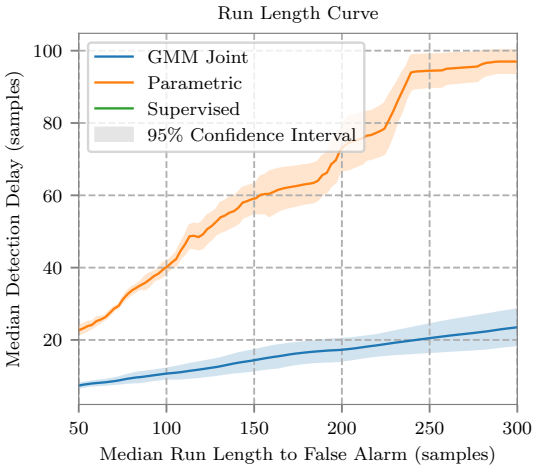
(A) ROC curves for the Limpopo data set band 6



(B) Detection Delay curves for the Limpopo data set band 6

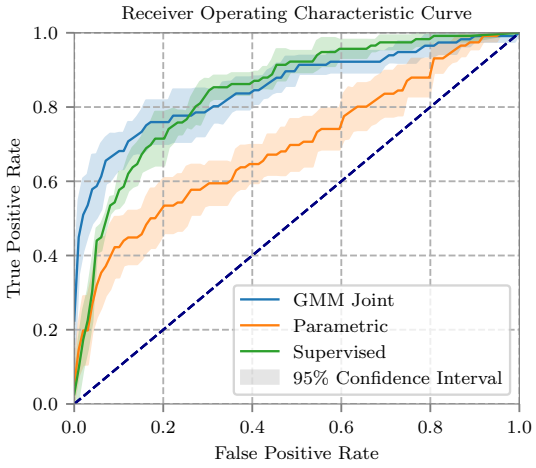


(C) ROC curves for the NSW data set band 6

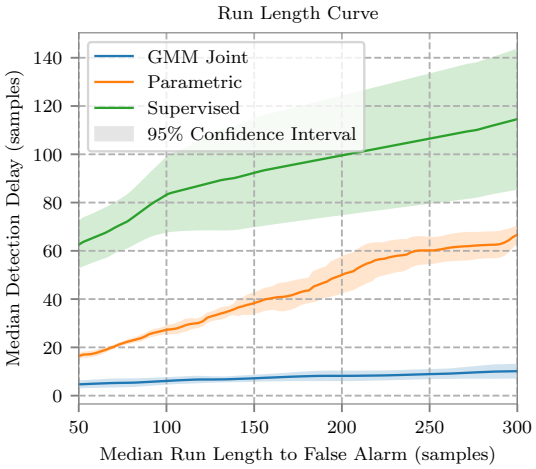


(D) Detection Delay curves for the NSW data set band 6

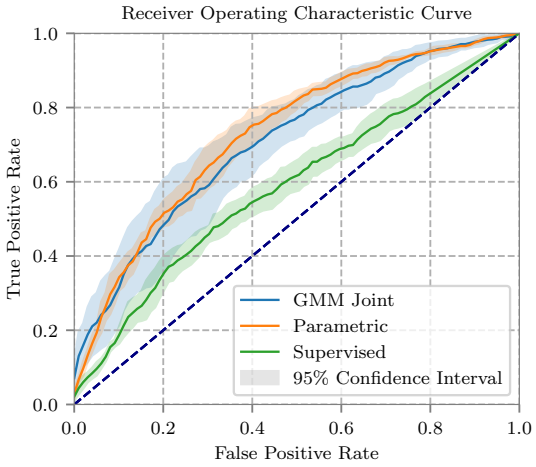
FIGURE B.6: ROC and Detection Delay curves for each of the considered methods using MODIS Band 6. Curves were evaluated with a look-back window size of $W = 100$ and best slack parameter.



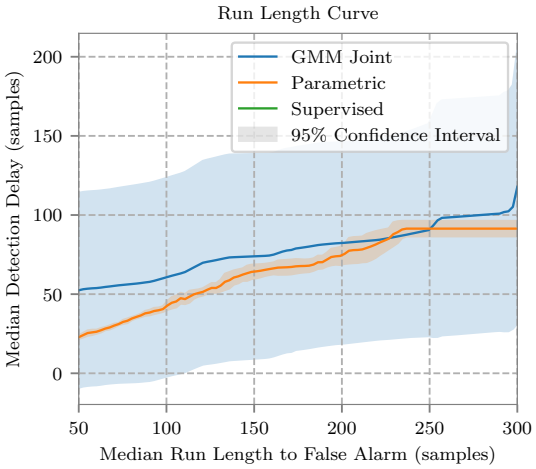
(A) ROC curves for the Limpopo data set band 7



(B) Detection Delay curves for the Limpopo data set band 7

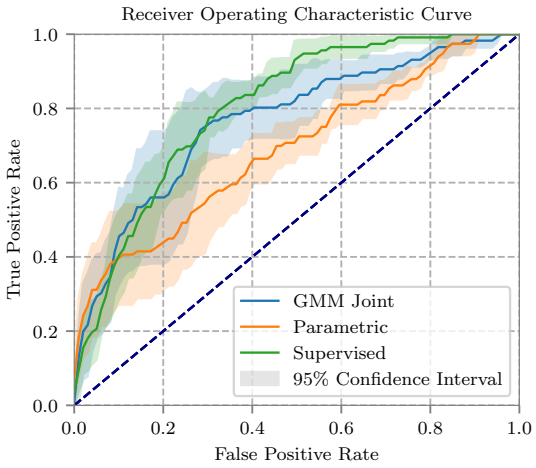


(C) ROC curves for the NSW data set band 7

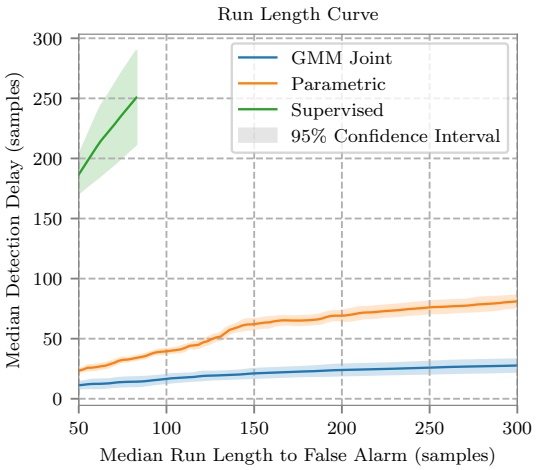


(D) Detection Delay curves for the NSW data set band 7

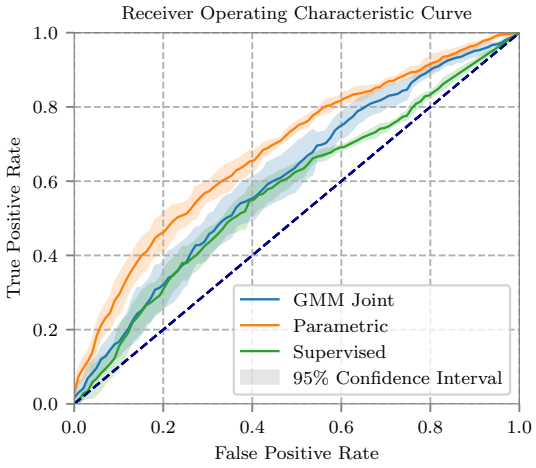
FIGURE B.7: ROC and Detection Delay curves for each of the considered methods using MODIS Band 7. Curves were evaluated with a look-back window size of $W = 100$ and best slack parameter.



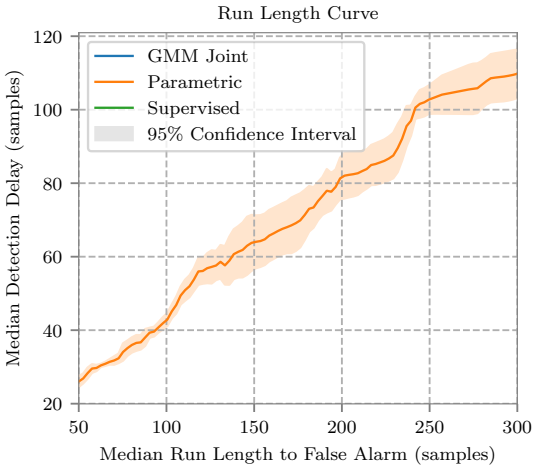
(A) ROC curves for the Limpopo data set NDVI



(B) Detection Delay curves for the Limpopo data set NDVI

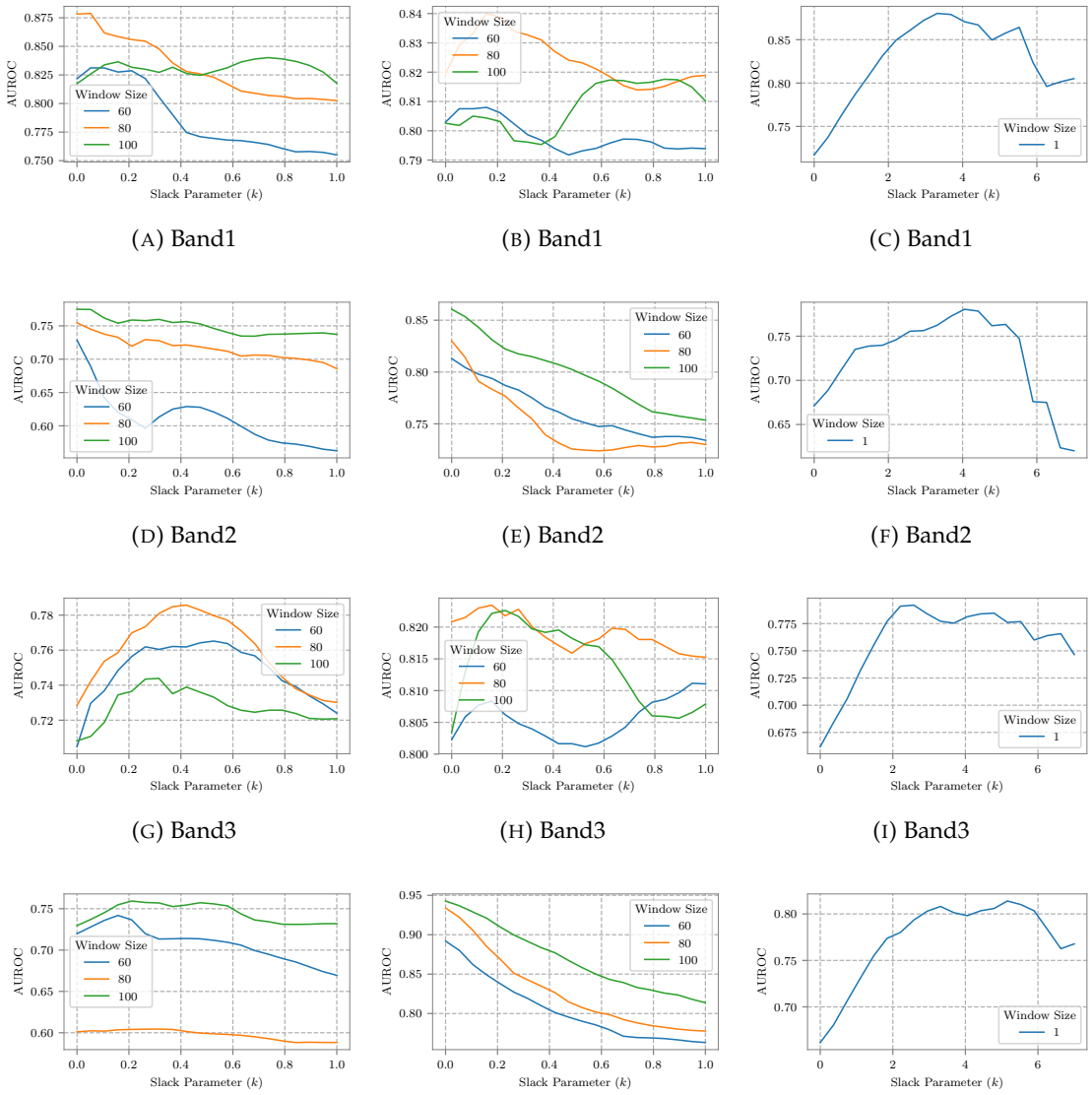


(C) ROC curves for the NSW data set NDVI



(D) Detection Delay curves for the NSW data set NDVI

FIGURE B.8: ROC and Detection Delay curves for each of the considered methods using MODIS NDVI. Curves were evaluated with a look-back window size of $W = 100$ and best slack parameter.



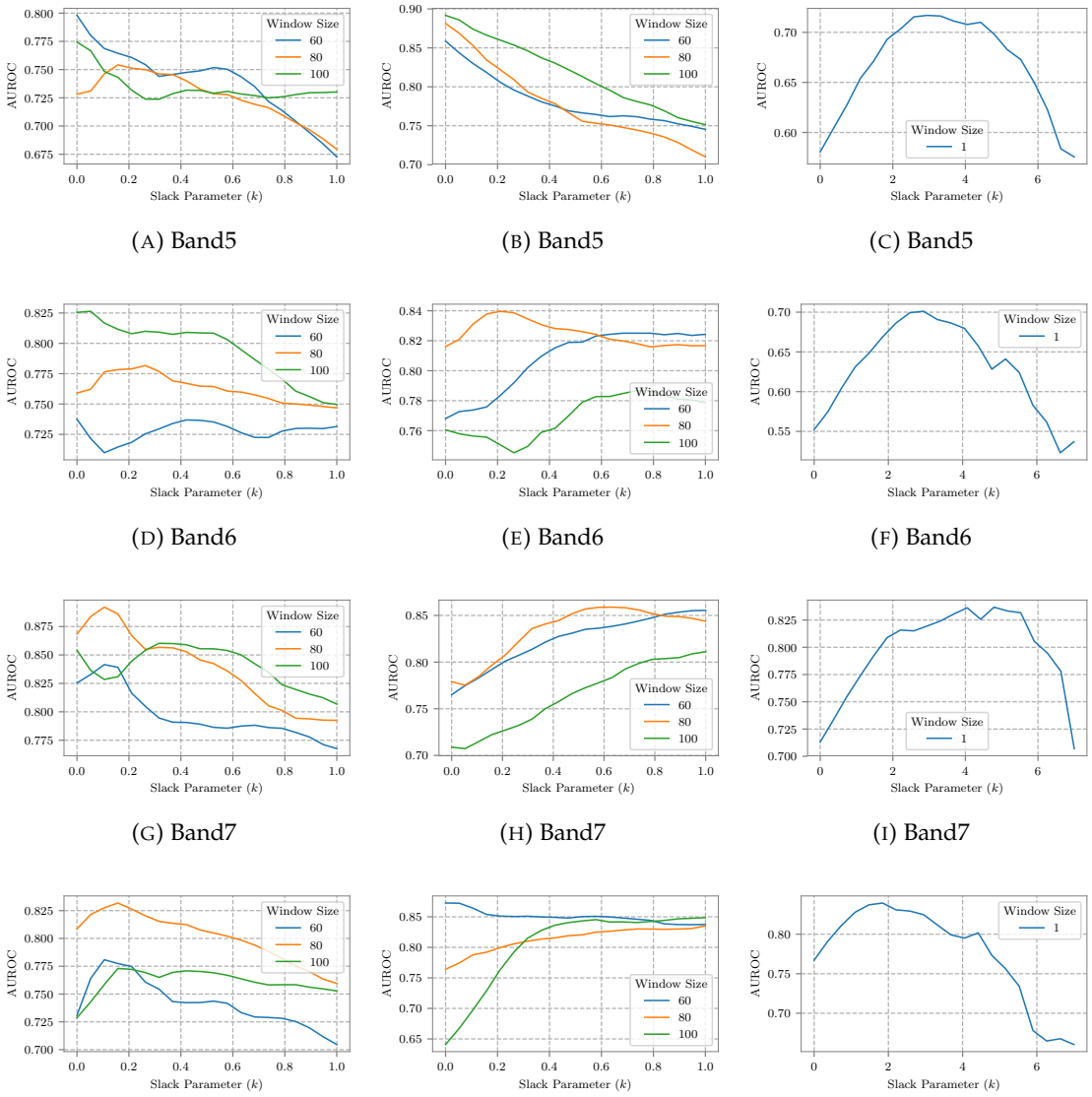
(A) Band1 (B) Band1 (C) Band1

(D) Band2 (E) Band2 (F) Band2

(G) Band3 (H) Band3 (I) Band3

(J) Band4 (K) Band4 (L) Band4

FIGURE B.9: Plots of AUROC as a function of CUSUM slack parameter, k, for the Limpopo data set. Columns left to right are the evaulated methods (GMM, Parametric, Supervised)



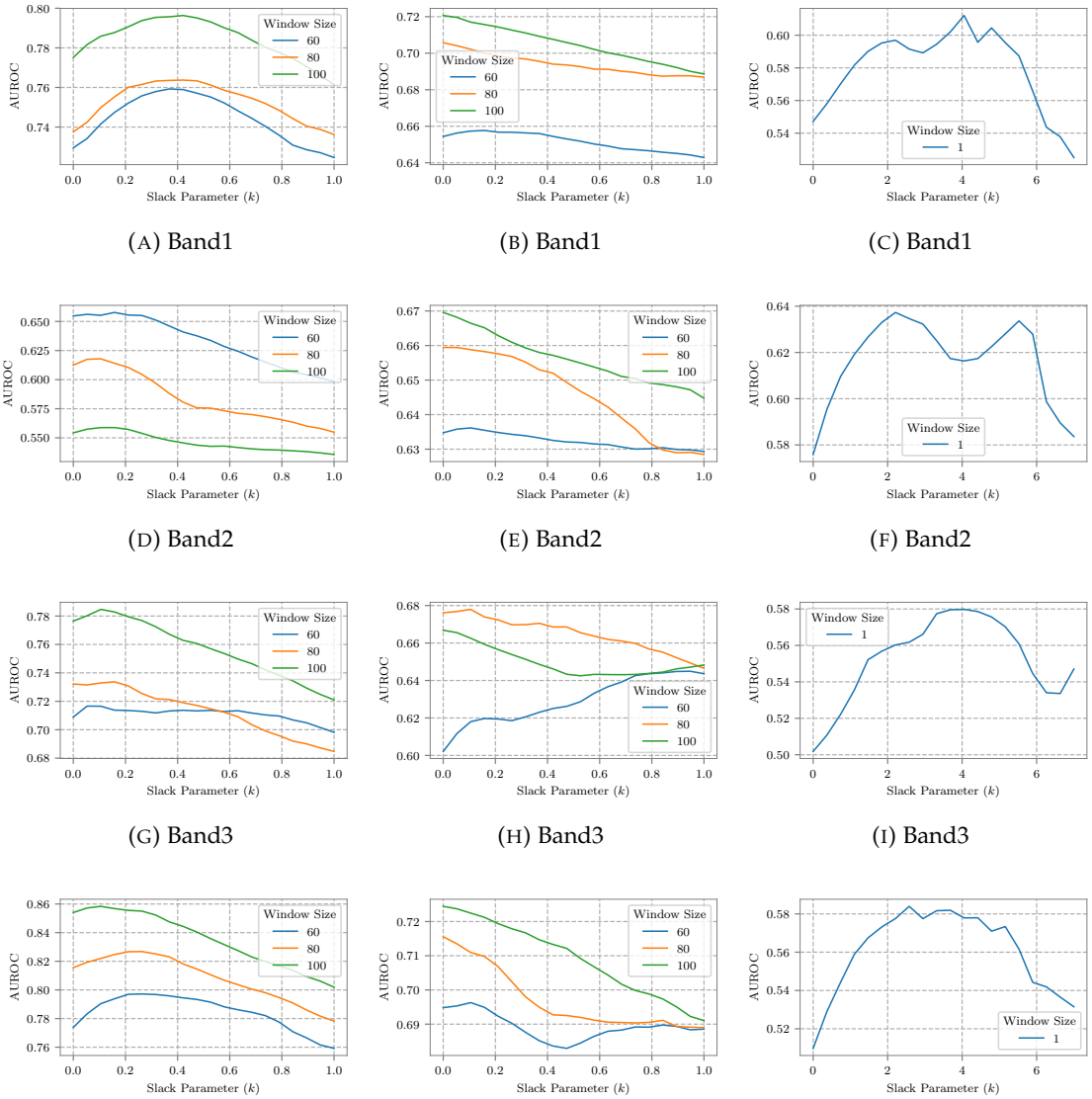
(A) Band5 (B) Band5 (C) Band5

(D) Band6 (E) Band6 (F) Band6

(G) Band7 (H) Band7 (I) Band7

(J) NDVI (K) NDVI (L) NDVI

FIGURE B.10: Plots of AUROC as a function of CUSUM slack parameter, k, for the Limpopo data set. Columns left to right are the evaulated methods (GMM, Parametric, Supervised)



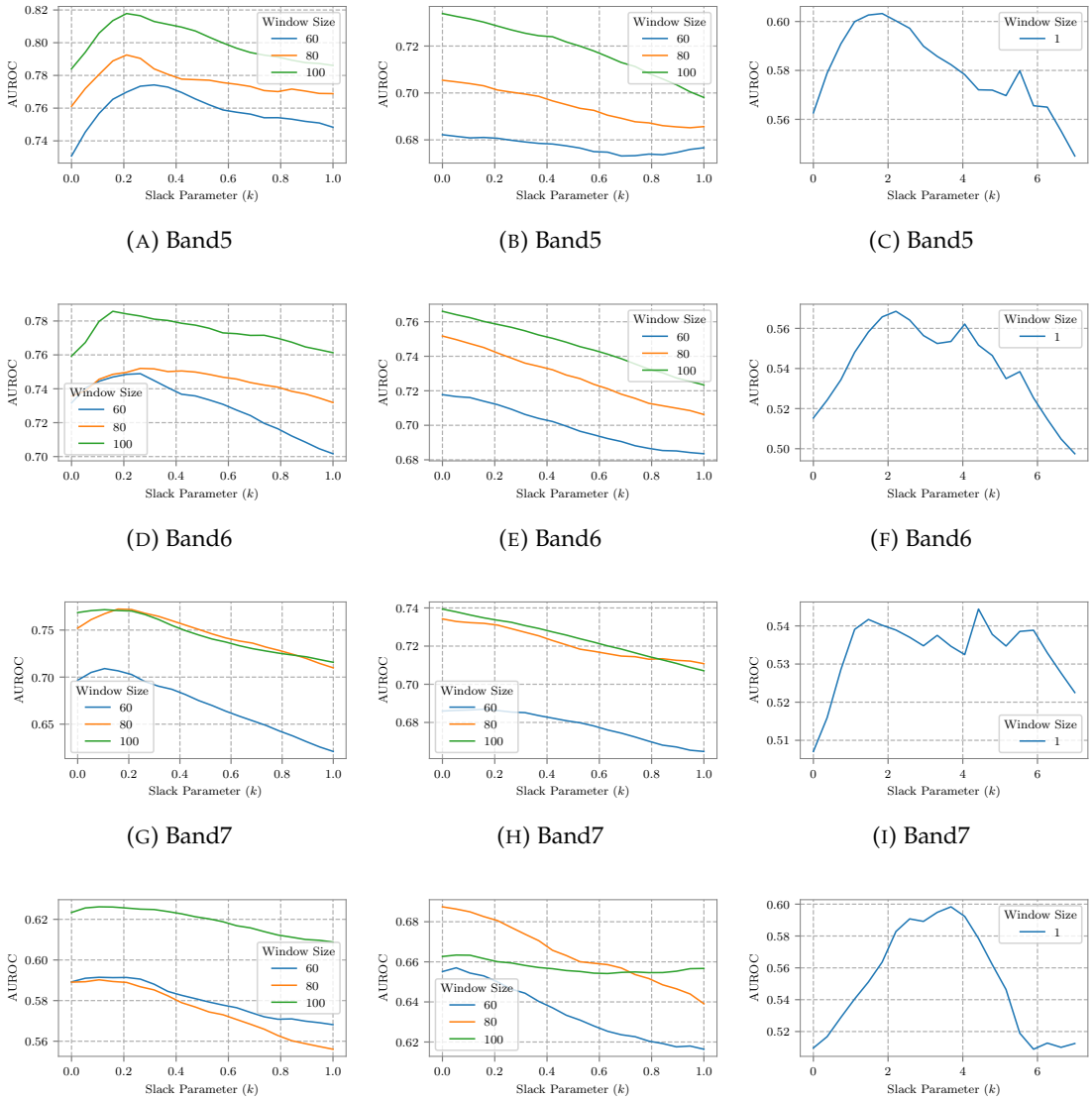
(A) Band1 (B) Band1 (C) Band1

(D) Band2 (E) Band2 (F) Band2

(G) Band3 (H) Band3 (I) Band3

(J) Band4 (K) Band4 (L) Band4

FIGURE B.11: Plots of AUROC as a function of CUSUM slack parameter, k, for the NSW data set. Columns left to right are the evaulated methods (GMM, Parametric, Supervised) and rows top to bottom are MODIS bands 1-4



(A) Band5 (B) Band5 (C) Band5

(D) Band6 (E) Band6 (F) Band6

(G) Band7 (H) Band7 (I) Band7

(J) NDVI (K) NDVI (L) NDVI

FIGURE B.12: Plots of AUROC as a function of CUSUM slack parameter, k, for the NSW data set. Columns left to right are the evaulated methods (GMM, Parametric, Supervised) and rows top to bottom are MODIS bands 5-7 and NDVI

Appendix C

Publications Eminating From This Thesis and Related Work

C.1 Publications from Thesis

- Olding, W. C., Olivier, J. C., Salmon, B. P., and Kleynhans, W. "Unsupervised Land Cover Change Estimation Using Region Covariance Estimates.", 2018 IEEE Geoscience and Remote Sensing Letters.
- Olding, W. C., Olivier, J. C., Salmon, B. P., and Kleynhans, W."A Forecasting Approach to Online Change Detection in Land Cover Time Series.", 2018 IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. (Accepted with revisions)

C.2 Related Work

C.2.1 Primary Author

- Olding, Willem C., Jan C. Olivier, and Brian P. Salmon. "A Markov random field model for decision level fusion of multi-source image segments." In Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International, pp. 2385-2388. IEEE, 2015.

C.2.2 Co-Author

- Salmon, Brian P., Waldo Kleynhans, Jan C. Olivier, Willem C. Olding, Konrad J. Wessels, and Frans Van den Bergh. "A modified temporal approach to meta-optimizing

an Extended Kalman Filter's parameters." In Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International, pp. 1144-1147. IEEE, 2014.

- Salmon, Brian P., Waldo Kleynhans, Jan C. Olivier, Colin P. Schwegmann, and Willem C. Olding. "A multi-tier higher order conditional random field for land cover classification of multi-temporal multi-spectral landsat imagery." In Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International, pp. 4372-4375. IEEE, 2015.

Appendix D

Publication for Offline Change Detection

Unsupervised Land Cover Change Estimation using Region Covariance Estimates

W.C. Olding, J.C. Olivier, B.P. Salmon, and W. Kleynhans

***Abstract*—In this paper we demonstrate the utility of estimating a probabilistic model of the underlying seasonal and inter-annual variations experienced by land cover time series in a given geographical region. Time series that deviate from these trajectories due to human induced change appear as outliers and can be detected using their Mahalanobis distance from the mean under the joint distribution of time samples. We apply this model to a collection of pixel time series acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS) platform over Limpopo Province, South Africa for the task of identifying human settlement expansion. For estimation of the time of change we present a hypothesis testing approach that tests for a decrease in correlation between samples before and after change. This was found to be highly effective, yielding a mean absolute error of 52 days.**

***Index Terms*—Change detection algorithms, Covariance matrices, Density estimation robust algorithm, Remote sensing, Time series analysis**

I. Introduction

REMOTELY sensed time series of vegetated land cover contain variations at both seasonal and inter-annual time scales [1]. Distinguishing between change in the underlying land cover and natural variations is a challenging problem requiring specialized methods [2]. Seasonal variability is approximately periodic and caused by vegetation phenology and reactions to annual changes in growing conditions such as rainfall, temperature and daylight length. Inter-annual variability is irregular and driven by climatic changes on a time scale greater than one year, for example periods of drought or above average rainfall driven by the El Niño Southern Oscillation (ENSO). Throughout this paper the combination of these effects will be referred to as the natural land cover dynamics. We define land cover change as significant and permanent variation driven by a factor other than the natural land cover dynamics, typically anthropogenic activities. Detecting change therefore requires a good estimate of the natural dynamics during the time span of interest. In this letter we also investigate the problem of identifying the time at which land cover change occurred. For certain applications, such as detection of settlement expansion, this information can be used to estimate the maturity of the settlement and the factors that caused the expansion to take place.

Early approaches to dealing with the seasonal variability in land cover time series involve aggregating data for each year and assessing the change between years [3]–[6]. This is effective for change detection but limits the temporal resolution at which the change points can be located.

Increasingly, methods are being proposed that consider the land cover time series without downsampling and apply techniques from the fields of regression, signal processing and time series analysis. In [2] a parametric model was used to decompose land cover time series into seasonal, trend and noise components. The trend component is modeled using piecewise linear regression and the seasonal component using piecewise regression of a simple harmonic model. In [7] an extended Kalman filter was used to track seasonal oscillations in MODIS time series. The inter-annual variations were modeled as a time variant bias component. The method by itself was unsupervised but required careful selection of the noise covariance matrices to perform well, a problem later addressed by [8]. The Extended Kalman filter has also been used for feature extraction in several subsequent methods [9], [10]. In [11] it was shown how a Gaussian Process with a modified periodic covariance function could be used to detect change in land cover time series by monitoring the difference between the observations and model forecast.

The presence of the vegetation dynamics generally results in time series that are non-stationary [12], a property that is assumed by classical change detection methods [13]. For example Page's Cumulative Sum algorithm (CUSUM) [14] assumes that the data is generated by a known, stationary, stochastic process and declares change as a statistically significant evolution in the outputs. Applying such methods for land cover change detection requires modification to account for the non-stationarity. [15] showed how CUSUM could be applied to MODIS time series with annual seasonal variations by using historical data to estimate independent densities for each sample time throughout the year. Inter-annual variations were not explicitly accounted for and were therefore considered by the model as noise.

[12] analyzed the correlation within MODIS time series by using the autocorrelation function as a feature extraction method to detect change. The autocorrelation was calculated as a function of lag under the wide-sense stationary assumption. The measured land cover time series were observed to be non-stationary due to inter-annual variation but experienced a much larger degree of non-stationarity when undergoing change.

We extend this concept further by estimating a Gaussian joint density over the signal time samples for the region of interest. This is fully defined by a mean signal, which captures the global trend, and a covariance matrix. No assumptions are made regarding the temporal stationarity of the data, only that the joint distribution of the samples in each length N signal

W.C. Olding, J.C. Olivier and B.P. Salmon are with the School of Engineering, University of Tasmania, Australia. (email: Willem.Olding@utas.edu.au)

W. Kleynhans is with the Remote Sensing Research Unit, Meraka Institute, CSIR, Pretoria 0081, South Africa.

is Gaussian.

The model is able to capture the underlying trends and the correlations between samples in the signal. The composition of the land cover within a pixel characterizes how the measured reflectance will respond to the natural driving forces both in bias and magnitude. For example a pixel composed of mostly bare soil may have a low reflectance in vegetation sensitive bands and respond very little to seasonal variation. A pixel composed of forest may have a higher average reflectance and also a greater response to seasonal variation [2]. Both this bias from the mean signal and the size of the response can be captured in the covariance matrix. By fitting a full N-dimensional Gaussian model it is possible to leverage correlation information to provide a better estimate of a sample's expected mean and reduce the variance.

Estimating the mean and covariance for the multivariate Gaussian model requires an unlabeled training set comprised of similar land cover types over the same time period and from the same geographical region. Time series from a similar geographical region may be different in their composition and response; however, we can assume that they all experience similar seasonal and inter-annual climatic changes.

This study focuses on off-line change detection and change point estimation. Off-line methods consider time series as static blocks of data and identify change within them. This is in contrast to on-line methods which accept streams of data and aim to identify change with minimal delay.

II. Study Area and Data

The time series used in this work were captured by the MODIS sensor over Limpopo Province, South Africa. Each pixel time series is comprised of a single MODIS band from the MDC43A4 product [16] over the time period of 2000-2012. This product provides surface reflectance values for bands 1-7 as well as the Normalized Difference Vegetation Index (NDVI) at 500m spatial resolution. The reflectance has been corrected to nadir at solar noon using a Bidirectional Reflectance Distribution Function (BRDF). The sample period is 8 days composited from a 16 day temporal acquisition window. For the time span considered in this study this yields $N = 381$ samples per time series. Samples containing significant cloud cover, as indicated by the MODIS product quality flags, were omitted and interpolated using a cubic spline.

The data set is comprised of MODIS pixel time series which have been labeled by a team of expert analysts as either *vegetation*, *rural settlement* or *change*. The study areas were randomly selected regions within the MODIS tile H20V11 which were mapped as rural settlements in a previous survey. The study areas and surrounding vegetation were manually classified using multi-date high resolution imagery. Change was declared if more than 70% of a pixel was converted from vegetation to rural settlement within the time span of the imagery. The number of time series of each class is as follows: *vegetation* - 997 (42%), *rural settlement* (53%) - 1235, *change* - 116 (5%). A comprehensive description of the data set can be found in [10]. As the method proposed in the paper is unsupervised, the time series labels were only used in the final stage to assess change detection performance.

III. Methodology

Let $\mathcal{X} = \{\mathbf{x}^{(i)}\}_{i=1}^{i=D}$ be a set of MODIS pixel time series for a single band captured over a geographical region. Each time series signal is a vector $\mathbf{x} = [x_1, \ldots, x_N]^T$. Each series can be considered an observation of a multivariate random variable $\mathbf{X} = [X_1, \ldots, X_N]^T$. Given a set of unlabeled time series the natural land cover dynamics can be modeled by fitting a set of univariate Gaussians, expressed as

$$X_k \sim \mathcal{N}(\mu_k, \sigma_k^2) \tag{1}$$

with the mean and variance estimated independently for each time sample. The probability of observing a time series under the independence assumption is given by

$$\Pr(\mathbf{X} = \mathbf{x}) = \prod_{k=1}^{N} \Pr(X_k = x_k). \tag{2}$$

Unless all of the signals are of the same homogeneous land cover, such a model will be dominated by the variance between signals and as a result will be insensitive to small deviations from the natural dynamics. To alleviate this we fit a single N-dimensional multivariate Gaussian model. Under this model the probability of observing a signal is

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{3}$$

with the mean and covariance matrix estimated from an unlabeled training set of time series from the region.

One further consideration is how to estimate the covariance matrix. The empirical method of covariance estimation

$$\boldsymbol{\Sigma} = \frac{1}{D-1} \sum_{i=1}^{D} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T \tag{4}$$

is notoriously sensitive to the presence of outliers in the data set. To alleviate this we estimate the covariance using the robust method of [17]. This method estimates the covariance using only h out of the total D examples. h examples are selected as the subset of samples that yields the covariance matrix with the smallest determinant. While the subset size, h, could be treated as a hyper-parameter we instead opted to use $h = \lceil \frac{D+N+1}{2} \rceil$. This is the minimum allowable value of h for which the algorithm will still converge correctly [17]. From our experiments we found this method to improve change detection performance compared with empirical estimates. The time complexity to estimate the covariance using the empirical and robust method is $\mathcal{O}(DN^2)$ and $\mathcal{O}(D^N)$ respectively [17].

A. Change Detection

For change detection we use an approach generally applied to outlier detection. Under this model, signal deviation from the natural dynamics manifests itself as an outlier to the distribution. Candidate signals are ranked according to their Mahalanobis distance to the mean, D_M, under the estimated multivariate Gaussian. A threshold, δ, applied to this distance

TABLE I
COMPARISON OF AVERAGE LOG-LIKELIHOODS ON SYNTHETIC CHANGE DATA

Method	Average Log-Likelihood
Direct Fit	1562
No-change Fit	1314
Modified Covariance	1457

is used to classify a signal as change or no-change. A signal, $\mathbf{x}$, is classified as having experienced change if

$$D_M(\mathbf{x}) = \sqrt{(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})} > \delta. \quad (5)$$

The value of the threshold, δ, may be used to trade between recall and false alarm rate. The time complexity for change detection is equivalent to that of calculating the Mahalanobis distance for each time series, $\mathcal{O}(DN^3)$.

B. Change Point Estimation

Consider a time series with a single land cover conversion at index k^*. We expect that values of the signal prior to k^* should be less correlated with values after k^* and visa versa as they are generated by different types of land cover. Furthermore we propose that a better fitting model can be obtained by assuming the values before and after k^* to be independent, allowing the joint distribution to be expressed as

$$\Pr_{k^*}(\mathbf{x}) = \Pr(x_1, \ldots, x_{k^*}) \Pr(x_{k^*+1}, \ldots, x_N). \quad (6)$$

Without refitting the model we can impose this assumption by zeroing the off-diagonal elements of the covariance matrix that relate values before and after the change point. In block matrix notation this can be expressed as

$$\boldsymbol{\Sigma}(k) = \begin{bmatrix} \boldsymbol{\Sigma}_{(1,\ldots,k^*)(1,\ldots,k^*)} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{(k^*+1,\ldots,N)(k^*+1,\ldots,N)} \end{bmatrix}. \quad (7)$$

In order to validate that the goodness-of-fit is indeed improved by this modification when change points are present, we compared it against the original model and also a Gaussian model fit directly from a set of change time series. A synthetic data set for this purpose was created by randomly sampling with replacement 10000 vegetation and rural settlement time series from the original data set and concatenating them with an identical and known change point at $k^* = 190$. Fig. 1 compares the covariance matrices estimated from the full data set (majority no-change) and the synthetic data set. It can be seen that the correlation between samples before and after k^* in the synthetic change signals are close to zero.

To quantify the goodness-of-fit we calculated the average log-likelihood for the synthetic data set. This gives the expected log probability of observing a time series from the data set under the given model. A higher value indicates a better fitting model. Tab. I shows that by enforcing independence post-fit the likelihood approaches that of the model fit directly to the change data.

When the time of change is unknown let us consider the hypothesis that the signal change point occurred at time index k, denoted as $\mathbf{H}_k$. Under this hypothesis we expect the likelihood to be improved by replacing the covariance of the

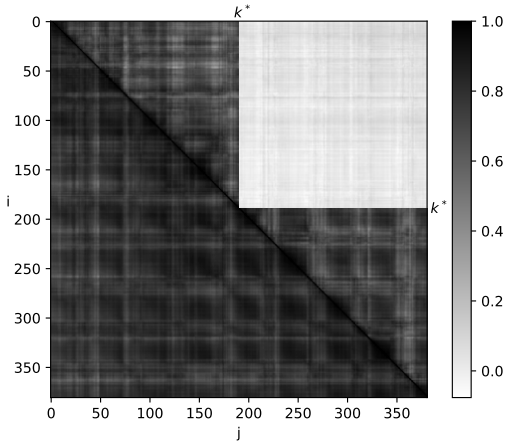


Fig. 1. Overlaid correlation matrix (normalized covariance matrix) heat maps estimated for MODIS band 1 over our region of interest. Lower diagonal is estimated from a set of no-change time series and upper diagonal is estimated from a set of synthetic change time series with change point $k^* = 190$.

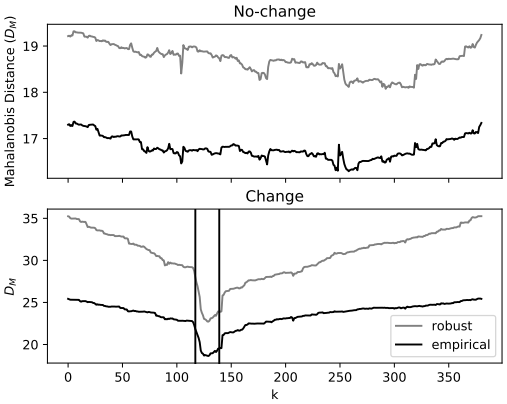


Fig. 2. Mahalanobis distance as a function of hypothesized change index, k. Vertical lines indicate the beginning and end of the synthetic change transition.

model with that of (7). Identifying the change point is then a matter of evaluating all possible hypotheses and selecting the one that maximizes the likelihood or equivalently minimizes the Mahalanobis distance. Under this model the best estimate of the change point is

$$\hat{k} = \underset{1 \le k \le N}{\arg\min} \sqrt{(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}(k)^{-1}(\mathbf{x}-\boldsymbol{\mu})}. \quad (8)$$

As the Mahalanobis distance must be recomputed at each candidate change point the time complexity in this case is $\mathcal{O}(DN^4)$. Fig. 2 illustrates how, for a synthetic time series containing a known change point, the model is an increasingly better fit as the proposed change point approaches the actual change point. It can be seen that for a no-change signal the distance does not change significantly.

As a comparison we also implemented a simple maximum-likelihood, supervised off-line change detector. This method fits a 1-dimensional Gaussian for each class (*vegetation* and *rural settlement*) at each time step, yielding a mean and

variance signal per class. If we define the probability that an observation was generated by a *vegetation* and *rural settlement* pixel at time step i as $\Pr_i^{(v)}(x_i)$ and $\Pr_i^{(r)}(x_i)$ respectively, we can define the maximum likelihood change point, k, as

$$\hat{k} = \underset{1 \leq k \leq N}{\arg\max} \prod_{i=1}^{k} \Pr_i^{(v)}(x_i) \prod_{i=k+1}^{N} \Pr_i^{(r)}(x_i). \quad (9)$$

This is valid as we know that only transitions from *vegetation* to *rural settlement* are present in the synthetic change signals.

IV. Experiment and Results

To evaluate both the change detection and change point estimation methods we applied them to each of the available MODIS bands as well as the NDVI. All 2348 time series were used as the unlabeled training set to estimate the mean and covariance matrix for the region using both the empirical and robust methods. For change detection the Mahalanobis distance was calculated for each time series using (5). We selected a value for the threshold, δ, using the labeled training set and evaluated the method using 10-fold cross validation. On each training set the threshold was selected to yield the best split by maximizing the information gain. To quantify the change detection performance we calculated the True Positive Rate (TPR) which gives the percentage of change signals correctly identified, False Positive Rate (FPR) which is the percentage of no-change signals incorrectly labeled and Accuracy (Acc), the percentage of examples correctly labeled. As the exact time of change was not known these were calculated based on detecting the presence of change at any point within the signal. Tab. II compares these performance metrics with other published work on this data set.

As the selection of the threshold, δ, is application dependent we also give results in terms of the Receiver Operating Characteristic (ROC) curve in Figure. 3a. This allows comparison between methods for any false positive or true positive rate. We also calculate the Area Under the ROC curve (AUROC) as a threshold independent performance metric.

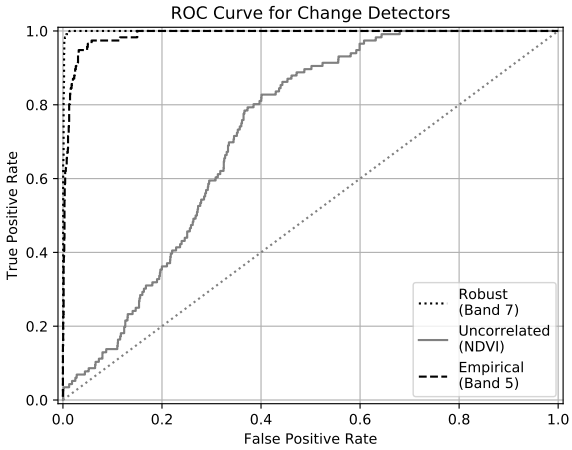
A synthetic set was created for evaluating the change point estimation methods. The synthetic data set was created using the same method as in [18]. Each synthetic time series is comprised of a randomly selected vegetation signal and a randomly selected rural settlement signal that are linearly blended over a period of six months. The change points are randomly selected from a uniform distribution spanning the length of the series with a six month buffer at the start and end. To evaluate the performance of the methods we consider the Mean Absolute Error (MAE) between the estimated change point and the known change point. As the synthetic change is a linear transition we arbitrarily define the actual change point as the center of the transition window. Tab. III gives the Mean Absolute Error (MAE) of the change point estimates for the synthetic data.

The average run times of our implementation for all 2348 time series on a standard desktop computer (Intel i7) were as follows: empirical train - 0.02s, robust train - 12.52s, change detection - 0.013s, change point estimation - 7.06s [1].

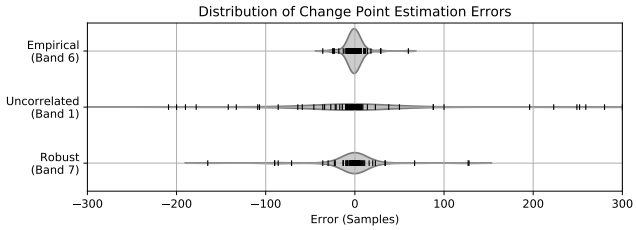
[1]Code implementation available at DOI: 10.5281/zenodo.1237328.

TABLE II
Comparison of performance metrics with other methods in the literature applied to the same data set. Bands with the highest TPR are shown.

Method	TPR (%)	FPR (%)	Acc (%)	AUROC
Robust Covariance (B7)	96.7	0.4	99.4	0.99
EKF + Non-linear Detector (7 bands) [10]	96	0.2	99	-
EKF + Non-linear Detector (1 band) [10]	96	1	99	-
Empirical Covariance (B1)	93.7	3.1	96.7	0.98
Extended Kalman Filter (EKF) [7]	89	13	87	-



(a) Receiver Operating Characteristic (ROC) curves for the evaluated change detection methods.



(b) Distribution of Errors for the evaluated change point estimation methods on synthetic data.

Fig. 3. Plots showing the relative performance for change detection and change point estimation. In each case the best performing bands for each method are presented.

TABLE III
Comparison of MAE between estimated and known change point on synthetic data. Bands with the smallest error are included.

Method	MAE (Samples)	MAE (Days)
Empirical (Band 6)	6.55	52.4
Robust (Band 7)	13.94	111.52
Supervised Uncorrelated (Band 1)	51.72	413.76

V. Discussion

The robust estimator applied to MODIS band 7 was found to be the most effective change detector although the performance of the method does not appear to be sensitive to the choice of the band (See Tab. IV). We hypothesize that it is the severe nature of the removal of vegetation and construction of

TABLE IV
COMPARISON OF PERFORMANCE METRICS FOR THE ROBUST ESTIMATE METHOD APPLIED TO DIFFERENT SPECTRAL BANDS

Band	TPR (%)	FPR (%)	Acc(%)	AUROC
1	93.1	1.0	98.7	0.998
2	95.8	3.1	96.8	0.994
3	96.6	1.0	98.9	0.997
4	90.5	1.3	98.2	0.993
5	94.8	1.8	98.0	0.996
6	96.4	1.4	98.5	0.997
7	96.7	0.45	96.7	0.999
NDVI	94.0	1.8	98.0	0.997

dwellings that results in a simultaneous change in reflectance across all bands. The superiority of the robust estimator suggests that the training set contained outliers and their presence degraded the change detection performance of the empirically fitted model. This improved detection performance is at the expense of increased computational complexity. This is seen in the significantly increased run time compared with the empirical method.

For change point estimation the error for the empirical covariance estimator shows superior performance to the robust model based on error distribution (Fig. 3b) and mean absolute error (Tab. III). This result is interesting as the robust estimator proved much more effective for change detection. This is likely due to the increased sensitivity of the robust model resulting in greater perturbations in the Mahalanobis distance vs change index curve. A significant perturbation may lead to an incorrect global minimum reducing the efficacy of the method. Fig. 2 appears to support this as the curve for the robust estimate shows a greater degree of deviation from the overall convex trend.

Even with access to significant domain knowledge, such as the fact that there are two classes and the transition in the synthetic data set is always from *vegetation* to *rural settlement*, the uncorrelated supervised model showed poor performance when compared to our unsupervised methods. This supports our hypothesis that the within signal correlations are a more informative feature for detecting change than the signal values alone.

VI. CONCLUSION

We showed how a multivariate Gaussian model fit from an unlabeled training set of geographically similar pixels can be effective for land cover change detection. Change signals appear as outliers with respect to the Mahalanobis distance under the fitted distribution. Robust estimates of the covariance were found to yield improved change detection performance. We also present a method for applying the same model to change point estimation by testing multiple hypotheses about the correlation between samples before and after the change point.

As the time series in this study were fairly short (381 samples) it was practical to fit the multivariate model of the entire signal. For very long time series the high dimensionality may adversely affect performance both in terms of accuracy and computation time. One solution is to limit the number of samples to a fixed window size and apply a sliding window approach to detect change. Such an approach may also enable the possibility of applying the method in an online setting allowing for real time applications.

REFERENCES

[1] X. Zhang, M. A. Friedl, C. B. Schaaf, A. H. Strahler, J. C. Hodges, F. Gao, B. C. Reed, and A. Huete, "Monitoring vegetation phenology using modis," *Remote Sens. Environ.*, vol. 84, no. 3, pp. 471–475, 2003.

[2] J. Verbesselt, R. Hyndman, A. Zeileis, and D. Culvenor, "Phenological change detection while accounting for abrupt and gradual trends in satellite image time series," *Remote Sens. Environ.*, vol. 114, no. 12, pp. 2970 – 2980, 2010.

[3] R. S. Lunetta, J. F. Knight, J. Ediriwickrema, J. G. Lyon, and L. D. Worthy, "Land-cover change detection using multi-temporal modis ndvi data," *Remote Sens. Environ.*, vol. 105, no. 2, pp. 142–154, 2006.

[4] S. Boriah, V. Kumar, M. Steinbach, C. Potter, and S. Klooster, "Land cover change detection: A case study," in *Proc. Int. Conf. Knowl. Discovery and Data Mining*, ser. KDD '08. New York, NY, USA: ACM, 2008, pp. 857–865.

[5] R. E. Kennedy, Z. Yang, and W. B. Cohen, "Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendrTemporal segmentation algorithms," *Remote Sens. Environ.*, vol. 114, no. 12, pp. 2897–2910, 2010.

[6] H. Yin, D. Pflugmacher, R. E. Kennedy, D. Sulla-Menashe, and P. Hostert, "Mapping annual land use and land cover changes using MODIS time series," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3421–3427, 2014.

[7] W. Kleynhans, J. C. Olivier, K. J. Wessels, B. P. Salmon, F. Van den Bergh, and K. Steenkamp, "Detecting land cover change using an extended kalman filter on MODIS NDVI time-series data," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 507–511, 2011.

[8] B. P. Salmon, W. Kleynhans, F. van den Bergh, J. C. Olivier, W. J. Marais, T. L. Grobler, and K. J. Wessels, "Meta-optimization of the extended kalman filters parameters through the use of the bias variance equilibrium point criterion," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5072–5087, 2014.

[9] B. P. Salmon, W. Kleynhans, F. van den Bergh, J. C. Olivier, T. L. Grobler, and K. J. Wessels, "Land cover change detection using the internal covariance matrix of the extended kalman filter over multiple spectral bands," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 3, pp. 1079–1085, 2013.

[10] B. Salmon, D. Holloway, W. Kleynhans, J. Olivier, and K. Wessels, "Applying model parameters as a driving force to a deterministic nonlinear system to detect land cover change," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7165–7176, 2017.

[11] V. Chandola and R. R. Vatsavai, "A Gaussian process based online change detection algorithm for monitoring periodic time series," in *Proc. Int. Conf. Data Mining*. SIAM, 2011, pp. 95–106.

[12] W. Kleynhans, B. P. Salmon, J. C. Olivier, F. van den Bergh, K. J. Wessels, T. L. Grobler, and K. C. Steenkamp, "Land cover change detection using autocorrelation analysis on modis time-series data: Detection of new human settlements in the gauteng province of south africa," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 3, pp. 777–783, 2012.

[13] S. Aminikhanghahi and D. J. Cook, "A survey of methods for time series change point detection," *Knowledge and information systems*, vol. 51, no. 2, pp. 339–367, 2017.

[14] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.

[15] T. L. Grobler, E. R. Ackermann, A. J. van Zyl, J. C. Olivier, W. Kleynhans, and B. P. Salmon, "Using page's cumulative sum test on MODIS time series to detect land-cover changes," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 332–336, 2013.

[16] C. Schaaf and Z. W, "MCD43A4 MODIS/Terra+Aqua BRDF/albedo nadir BRDF adjusted L3 global - 500m v006." in *NASA EOSDIS Land Processes DAAC*, 2015.

[17] P. J. Rousseeuw and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, 1999.

[18] W. Kleynhans, B. P. Salmon, J. C. Olivier, F. van den Bergh, K. J. Wessels, and T. Grobler, "Detecting land cover change using a sliding window temporal autocorrelation approach," in *Proc. IEEE Int. Geosci. Remote Sens. Symp*, 2012, pp. 6765–6768.

Appendix E

Publication for Online Change Detection

A Forecasting Approach to Online Change Detection in Land Cover Time Series

W.C. Olding, J.C. Olivier, B.P. Salmon, and W. Kleynhans

***Abstract*—We present a method for online detection of land cover change based on remotely sensed time series. Change is detected by monitoring deviations between observations and forecasts made using the time series historical data and similar time series in the geographical region. This method and several others were applied to MODIS 8-day surface reflectance data for problems of detecting settlement expansion in Limpopo Province, South Africa and detecting deforestation in New South Wales, Australia. The proposed method had significantly shorter median detection delay for equivalent rates of false alarms compared with the other evaluated methods. We obtained a median detection delay of 7 samples for settlement detection and 14 samples for deforestation detection corresponding to 56 days and 112 days respectively. This is compared with a median detection delay of 224 and 544 days for the best other methods evaluated. We suggest that the proposed method is an excellent candidate for land cover change detection where rapid detection is essential.**

***Index Terms*—Change Detection Algorithms, Remote Sensing, Sequential Detection, Time Series Analysis**

I. Introduction

One of the major goals of remote sensing satellite systems is to enable large scale monitoring of the earths land cover. Detecting when and where land cover change has occurred forms an important part of the monitoring process. Change detection is typically used in one of three cases. It may be used as the first stage in a processing pipeline, for example when updating land cover maps it can reduce the problem to only reclassifying regions which have changed. It may also be applied to produce useful information in its own right, for example estimating how much of a particular region has experienced change within a given time span. The third case, which will be the focus of this paper, involves monitoring land cover in near real-time in order to detect change shortly after it has occurred. In this case the output of a method can be used to guide action in response to the change.

Pixel time series generated by repeated observations from high temporal resolution multi-spectral satellite platforms have been shown to be well suited to detecting changes in vegetated land cover [1]–[9]. Detection of change from pixel reflectance time series is made difficult by several factors. Difficult to predict inter-annual variations coupled with semi-regular seasonal oscillations in reflectance make the time series highly non-stationary. This makes detecting land cover change a difficult problem requiring specialized methods [2]. Secondly, exactly what constitutes land cover change can be difficult to define. Within this work the focus is on detecting human induced changes that take place over a 6-12 month period. This is typical of the types of land cover change associated with gradual clearing of forest or urban expansion.

One advantage of remotely sensed time series is availability of many similar land cover time series within a region of interest. This collection of time series can be used to estimate the natural variation in reflectance caused by the influence of climatic changes on vegetation phenology. In this paper we propose an online change detection method that used this technique to detect deviations of a time series from its natural variations implying a change in the underlying land cover.

Common themes in existing land cover time series classification methods involve transformation of the time series into a form that is approximately stationary under no-change conditions. Classical change detection methods that assume stationarity can then be applied [10]. Some approaches assume the original time series is cyclo-stationary [6] while others make assumptions regarding the structure of the trend and periodicity [3], [7], [8]. These types of methods can be further split into supervised methods, which require a labeled training set to estimate class specific statistics [6], [9], and unsupervised methods which have no such requirement [3], [5], [7], [8].

In [6] change is detected from one known class to another by estimating class specific densities and calculating the likelihood ratio. The log-likelihood ratio was monitored for change using Page's Cumulative Sum (CUSUM) algorithm. To account for seasonal variation a density was calculated for each class for each time step throughout the year. Different years were aggregated so inter-annual variation was not explicitly accounted for.

A supervised approach is presented in [9]. They propose training a probabilistic classifier (Random Forest) at each time sample to estimate time series of class membership probabilities. These are monitored for change using the MODTrendr Algorithm [4]. As a supervised method this requires a large number of labeled no-change time series to fit the probabilistic classifiers. It also results in a large number of time series to monitor for change if there are many classes being considered.

[5] details a method designed specifically for detecting change in periodic, non-stationary time series, not limited to land cover. It operates by using Gaussian process regression to make a probabilistic forecast of the next sample. Once the next sample is available its z-score under the forecast distribution is calculated. By repeatedly applying this forecasting method a time series of z-scores is produced. Given the model is working correctly the resulting time series of z-scores should be

W.C. Olding, J.C. Olivier and B.P. Salmon are with the School of Engineering, University of Tasmania, Australia. (email: Willem.Olding@utas.edu.au)

W. Kleynhans is with the Remote Sensing Research Unit, Meraka Institute, CSIR, Pretoria 0081, South Africa.

Gaussian Independent Identically Distributed (IID). A series of samples that deviate significantly from the forecast will result in a change in distribution of the series of z-scores. This is detected using an Exponentially Weighted Moving Average (EWMA) control chart. The kernel for the Gaussian process was a periodic exponential allowing for modeling of periodic functions. The parameters for the kernel were fit at the start of the signal and then held constant. The computational performance and scalability of the method is improved upon in [11].

In [3] a parametric model comprised of a bias, linear trend and periodic seasonal component is used to make predictions. Three periodic terms are included in the model, a fundamental frequency equal to one cycle per year and two additional harmonics. This model is extrapolated into the future and the moving sum (MOSUM) of residuals over a finite lookahead window is monitored for deviation from the fitted model. The probability of observing a particular MOSUM with respect to the time series noise level is derived and a change alarm is triggered when it exceeds some significance level.

A parametric model based method is also applied in [7]. The authors take the approach of fitting a triply-modulated cosine and bias to land cover time series over a sliding window. The time series of parameters from each sliding window step is monitored for change. It is suggested that the bias parameter of the model should remain constant during a period of no-change and deviate when the land cover experiences change. The bias parameters series is converted to a series of z-scores by estimating parameters from its historical values. Change is declared when 6 out of 10 previous z-scores exceed some threshold.

[8] applies a parametric model identical to that of [3]. In this case residuals from the model over the entire series are converted to z-scores and this time series is directly monitored for outliers. Z-scores exceeding a given significance level are considered indicative of a change point. While this method uses a similar approach to those above it cannot be considered as an online approach as it requires the entire data set data in order to fit the parametric model and thus can only discover changes in the past.

The method we propose combines the ideas of [5] and [3] of using the prior samples of a time series to make a forecast with ideas of [6] and [9] of considering samples of other time series in the same region to estimate a baseline from which to detect change. Under our unified framework a collection of similar time series from the same region are used to estimate the joint distribution of the time series over a temporal window. This is conditioned on previous observations of a signal in order to make a prediction and calculate z-scores. The time series of z-scores is then monitored for change.

II. Background

A. Stationary of Land Cover Time Series

A stochastic process can be said to be *strictly stationary* if the joint distribution of any combination of values is time invariant. For a strictly stationary process, $\{X_t\}$, we can say for any set of time indices $\{t_1, \ldots, t_k\}$, the joint distribution

$$\Pr(X_{t_1+\tau}, \ldots X_{t_k+\tau}) \tag{1}$$

is identical regardless of the time shift τ. A special case of stationary is IID. As the name suggests each sample is independently drawn from the same distribution. In this case the joint distribution may be factorized as

$$\Pr(X_{t_1+\tau}, \ldots X_{t_k+\tau}) = \prod_{i=1}^{k} \Pr(X_{t_i+\tau}). \tag{2}$$

A process is said to be *cyclo-stationary* if the process obtained by sampling at integer multiples of some period T exhibits stationarity. Formally

$$\Pr(X_{t_1+T+\tau}, \ldots X_{t_1+nT+\tau}), n \in \mathcal{Z} \tag{3}$$

remains identical for any value of t_1 and τ and for some period T.

Several factors combine to make remotely sensed land cover time series fail to meet these stationarity assumptions. Most land cover is vegetated and this results in the dominant source of non-stationarity in the signal. It is well documented that the reflectance of vegetation varies approximately periodically in a yearly cycle [1]. In many cases this may result in cyclo-stationarity however the magnitude and bias of the oscillations is often non-constant between years due to inter-annual climatic changes such as droughts. In many cases trend components are present in vegetated time series that suggest either long term non-periodic change or change on a cycle much longer than the time span of available data.

III. Change Detection by Forecasting

Let us consider an estimator that is able to produce forecasts of the next value in a time series under no-change conditions. We denote the forecast at time t as $\hat{x}_t$. The residual of the forecast is defined as $r_t = x_t - \hat{x}_t$. A sufficiently good estimator should yield a series of residuals which are uncorrelated, unbiased (i.e. zero mean) and Gaussian. If the estimated residuals also have constant variance we can say the resulting time series of residuals is IID. Furthermore, if the residuals do not have constant variance but the estimator is able to produce good estimates of the variance of each sample, σ_t^2, the residuals can be converted to z-scores by

$$z_t = \frac{r_t}{\sigma_t} \tag{4}$$

and the series of z-scores will be IID unit Gaussian.

If the original time series contains a change point that is unforeseen to the estimator this will induce a bias in the series of z-scores/residuals which can be detected by a classical change detection method. An ideal estimator should yield a series of z-scores that is Gaussian IID under normal circumstances and respond strongly and rapidly in the presence of a change point. Such an estimator may take advantage of historical data, neighboring pixel time series or ancillary data in order to make its prediction.

One such method that follows this general framework is the GPChange method of [5], [11]. In this case the prediction is made using only historical data from the same signal

$$\Pr(X_t) = \Pr(X_t \mid X_{t-1}, X_{t-2}, \dots X_0). \tag{5}$$

Motivated by these previous works we propose a method that takes advantage of both the previous samples in the time series and also the time series in the surrounding region when making a prediction. This is done by directly estimating the joint distribution over a temporal window using a set of similar time series. As the model is re-estimated at each time step this does not require any assumptions of stationarity. Furthermore it does not impose assumptions on the structure of the data as in [3], [7], [8].

A. Proposed Method

We propose modeling the joint distribution over the samples in a temporal window as a multivariate Gaussian. Such a model can capture the mean signal of the region of interest, the uncertainty associated with the mean signal and the correlations between samples in the time series. The resulting model can be considered in a similar sense to a Gaussian process although discrete and with covariance estimated directly from data rather than predefined by a covariance function. By estimating the joint distribution from nearby time series it is possible to capture the natural variations of a region. This statistical information can be leveraged to detect when a time series moves in a way that is inconsistent with both its own historical trajectory and that of the surrounding region.

With access to a fitted joint distribution it is possible to condition on the previous observations in a signal to obtain a univariate distribution over the last sample. More explicitly let $\boldsymbol{x} = [x_t, \dots, x_{t-W-1}]^T$ be a vector of time series observations in a temporal window of size W. Let $\mathcal{X} = \boldsymbol{x}^{(i)}{}_{i=1}^{i=D}$ be a set of D time series from a similar geographical region and ideally containing similar land cover. We can assume that each of these time series windows is an observation of a multivariate random variable $\boldsymbol{X} = [X_t, \dots, X_{t-W-1}]^T \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Under the multivariate Gaussian assumption we can calculate the mean vector and covariance matrix as

$$\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}] \tag{6}$$

and

$$\boldsymbol{\Sigma} = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] \tag{7}$$

which can be estimated directly from the data.

For a single time series, $\mathbf{x}^{(i)}$ this model can be conditioned to calculate the univariate distribution over the last time sample given earlier observations within some temporal window, $\Pr(x_t^{(i)} \mid \mathbf{x}_{obs}^{(i)})$, using the standard method for conditioning a multivariate Gaussian. If we partition the covariance matrix into observed and unobserved variables as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_t^2 & \boldsymbol{\sigma}_{1,2} \\ \boldsymbol{\sigma}_{2,1} & \boldsymbol{\Sigma}_{2,2} \end{bmatrix} \tag{8}$$

where

$$\boldsymbol{\sigma}_{1,2} = \boldsymbol{\sigma}_{2,1}^T = [\mathrm{cov}(X_t, X_{t-1}), \dots, \mathrm{cov}(X_t, X_{t-W-1})] \tag{9}$$

and

$$\boldsymbol{\Sigma}_{2,2} = \begin{bmatrix} \sigma_{t-1}^2 & \dots & \mathrm{cov}(X_{t-1}, X_{t-W-1}) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(X_{t-W-1}, X_{t-1}) & \dots & \sigma_{t-W-1}^2 \end{bmatrix} \tag{10}$$

the conditional distribution is univariate Gaussian with mean and variance given by

$$\bar{\mu}_t^{(i)} = \mu_t + \boldsymbol{\sigma}_{1,2}\boldsymbol{\Sigma}_{2,2}^{-1}(\mathbf{x}_{obs}^{(i)} - \mu_{obs}) \tag{11}$$

and

$$\bar{\sigma}_t^2 = \sigma_t^2 - \boldsymbol{\sigma}_{1,2}\boldsymbol{\Sigma}_{2,2}^{-1}\boldsymbol{\sigma}_{2,1}. \tag{12}$$

From this probabilistic forecast the z-score of the observation of the i^{th} time series can be calculated as

$$z_t^{(i)} = \frac{x_t^{(i)} - \bar{\mu}_t^{(i)}}{\bar{\sigma}_t^{2(i)}}. \tag{13}$$

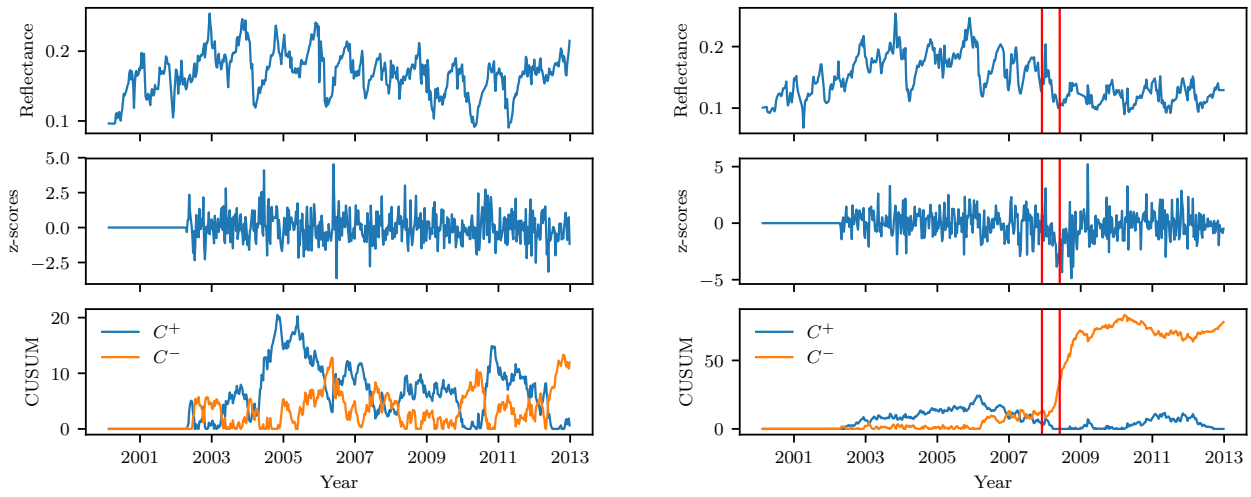
Repeatedly applying this method as the sliding window is shifted results in a time series of z-scores. The time series of z-scores can then be monitored for persistent deviation from the standard normal which implies a change.

If the region contains a very diverse land cover classes it might be the case that the variance of the fitted multivariate Gaussian is too high to produce useful forecasts. We suggest two alternative methods for estimating the joint Gaussian density. The choice of method depends on domain knowledge of the land cover types in the region of interest. In the case that there is one single dominant land cover type that should be monitored for change the robust estimator of [12] is well suited. This method estimates the mean and covariance matrix using a fixed size subset of the samples. The subsample is selected from all possible combinations as the one that yields the covariance matrix with the smallest determinant. This is very effective at eliminating outlying time series in the data set.

Another possibility is that there are multiple types of land cover type present in the region of interest. In this case a Gaussian Mixture Model approach can be used to simultaneously estimate multiple Gaussian joint densities and group similar time series. The joint distribution of the set of time series is then expressed as

$$\Pr(\mathbf{X}) = \sum_{c=1}^{k} w_c \mathcal{N}(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c). \tag{14}$$

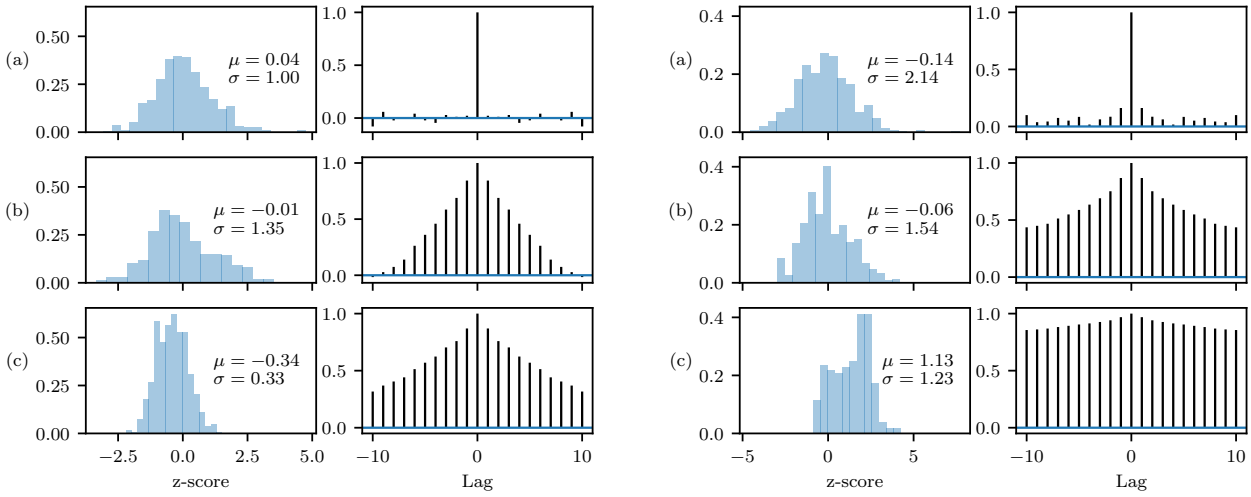
The weights, w_i, and parameters of the Gaussian distributions are estimated using the Expectation Maximization (EM) algorithm. The number of components must be selected in advance using location specific knowledge. When making a forecast the cluster that a time series belongs to can be estimated by finding closest component mean using the samples in the window. The mean vector and covariance matrix from this component are then used in equations (11) and (12). This is the variation we focus on in this study.



(a) No change time series

(b) Change time series

Fig. 1. Reflectance time series, z-score series and CUSUM output for a no-change series and a synthetic change from the Limpopo data set. Vertical bars indicate the beginning and end of the linear transition.



(a) No change time series

(b) Change time series

Fig. 2. Density plot and autocorrelation plot for each of the z-score extraction methods investigated in this paper. (a) Joint (GMM Method), (b) Parametric, (c) Univariate Supervised. Evaluated on a no-change time series and a synthetic change time series from the Limpopo data set using MODIS Band 1.

B. Change Detection on Z-Scores

Once the time series of z-scores has been produced by one of the methods discussed above it can be monitored to detect deviations from its expected distribution. Provided there are no changes present in the time series and the forecasting algorithm is working correctly the z-score time series should be IID Gaussian. To detect statistically significant persistent deviations from this distribution we apply Page's CUmulative SUM (CUSUM) algorithm [10], [13].

We present the CUSUM method under the assumption that the data is IID Gaussian with known mean and variance prior to change, after which it experiences a change in mean of unknown magnitude.

Let $\{z_i\}_{i=0}^{i=N}$ be the series of z-scores. By definition this series should be IID Gaussian with zero mean and unit variance. We therefore expect under normal conditions that the cumulative sum

$$C_i = \sum_{j=1}^{i} z_i \tag{15}$$

$$= z_i + C_{i-1} \tag{16}$$

is a random walk with zero mean [10]. If there is a persistent

bias in the models ability to make predictions, such as the underlying data generating distribution changing, we expect the random walk to drift away from zero in the direction of the bias. To improve the robustness of the method a slack parameter, k, is included [10]. This parameter is subtracted from the CUSUM at each iteration and reduces the sensitivity to small mean shifts which can help prevent false alarms. The cumulative sums are also split into positive and negative components with each clamped at zero. The resulting cumulative sums are defined recursively as

$$C_i^+ = \max[0, z_i - k + C_{i-1}^+] \quad (17)$$

$$C_i^- = \max[0, -z_i - k + C_{i-1}^-]. \quad (18)$$

[10]

As we do not know how the predictor will respond to a land cover change we monitor both cumulative sums. When either C^+ or C^- exceeds some threshold λ a change alarm is triggered and the corresponding cumulative sum is reset to zero. Fig. 1 shows the reflectance time series, z-score series extracted using the joint estimation method and the corresponding CUSUMs. In the time series containing a change it is possible to observe a negative bias in the z-score series immediately after the change point. This causes a strong response in C^-. In the no-change series the corresponding z-score series should be close to IID Gaussian noise and the CUSUMs of this signal will not significantly deviate from zero.

It can be shown that if the mean and variance before and after the change point is known and the change takes place instantaneously this method is optimal for a given detection delay [14]. This does not hold in our case as the distribution of the z-scores post change cannot be known.

IV. Study Areas and Data

A. Study Areas

We evaluate the methods by applying them to two applications that require rapid and accurate change detection.

The first problem is that of detecting settlement expansion in the Limpopo Province of South Africa. In some areas of this province settlement expansion is carried out ad-hoc by residents without prior planning approval. The local native vegetation of this region is predominately Savannah, characterized by small sparse trees and grassland [15]. Construction of settlements is characterized by the clearing of native vegetation and the construction of small dwellings [16]. Detection of unplanned settlements in this region has been the focus of several studies [6], [17]–[19].

The second problem relates to detecting changes in areas of native forest in New South Wales, Australia. Under the Australian government Emissions Reduction Fund (ERF) initiative, land owners are able to generate Australian carbon credit units (ACCUs) in exchange for preventing deforestation of native forest for which a clearing permit has previously been issued. To be eligible the native forest must have at least 20% canopy coverage with tree height greater than two meters [20].

At the commencement of a project an extensive audit is undertaken by an accredited third party and estimates made of the appropriate number of credits to be allocated throughout the project. A number of subsequent audits are also required to ensure compliance throughout the project lifetime. Projects have a permanence period of either 25 or 100 years. During this time period no clearing, with the exception of minimal thinning ($< 5\%$), is permitted [20].

At the time of writing there are approximately 400 vegetation projects underway in Australia. A system of automated change alarms based on remotely sensed time series has the potential to significantly reduce the auditing workload and target it to locations where the forest cover has changed and the project may require reassessment. Continuous monitoring is also important to guarantee the integrity of the carbon credit units. We apply online change detection methods to several regions in rural New South Wales (NSW) which are currently generating carbon credits in exchange for deforestation prevention.

B. Data Preparation

For both applications we make use of time series from the MODIS MCD43A4 product [21]. This product delivers Bi-Directional Reflectance Distribution Function (BRDF) corrected surface reflectance data from both the Terra and Aqua Platforms. The temporal resolution is 8 days with each data point selected as the highest quality acquisition in a 16 day temporal window. The spatial resolution is approximately 500m. This product has been used in numerous land cover change detection studies mostly due to its high temporal resolution and large catalog of data.

The data set for the Limpopo region is comprised of $D = 2348$ MODIS pixel time series over the time span of the years 2000 to 2013. Each was classified by an expert analyst as *vegetation* - 997 (42%), *settlement* (53%) - 1235, *change* - 116 (5%) by visual inspection of high resolution imagery. A pixel was labeled as change if it transitioned from vegetation to settlement within the time span of interest. The exact time of change is not known. A comprehensive description of the data set can be found in [19]. This type of change is classified as a land cover conversion. To quantify the detection delay of the methods in this study it was necessary to create synthetic change time series where the exact time of change can be controlled. To generate synthetic change time series we follow a similar approach to [18]. Two time series of different land cover types are linearly blended over a period of six months. This slow transition between land cover types aims to imitate the gradual conversion that is expected in the construction of a settlement.

The data set for New South Wales (NSW) is comprised of $D = 1994$ MODIS pixel time series located within regions assigned to a project which is currently earning carbon credits in exchange for avoided deforestation. These time series span the years of 2008 to 2018. Each time series has a corresponding pair which is closely located and contains a similar type of land cover but at a lower density. As no real change was observed within the designated time span only synthesized change series

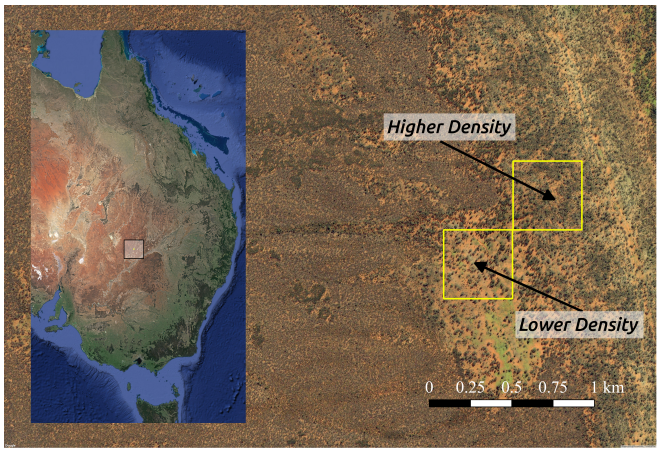


Fig. 3. The study region in New South Wales. This includes an example of two MODIS pixels within a protected area that contain similar land cover with differing densities. This pair of pixel time series is blended together over a 6 month transition to simulate the partial and gradual clearing of vegetation. (Image courtesy of Google Earth, DigitalGlobe.)

were considered. These change series were created by linearly blending over a period of 6 months between a pixel and its lower density pair. This aims to simulate the partial and gradual clearing of vegetation. Fig. 3 shows the study region and an example of two nearby pixels of differing densities. This problem is an example of within-class change detection [1].

The remotely sensed time series for both of these locations contained only a few missing values ($< 0.1\%$). The missing values present were filled using linear interpolation.

We consider only MODIS band 1 (620nm-670nm) and band 2 (841nm-876nm). These have been shown in previous studies [6], [19] to be most effective at detecting changes in vegetation. We also consider the Normalized Difference Vegetation Index (NDVI), a commonly used quantity defined for MODIS as the difference between bands 2 and 1 divided by their sum.

V. Experimental Methodology and Results

A. Evaluated Methods

We compare three methods for extracting time series of z-scores that are ideally IID under no-change conditions. Our proposed method of directly estimating the joint distribution using the Gaussian Mixture Model (Joint), a supervised univariate density estimation method similar to that of [6] (Univariate) and a parametric forecasting method similar to that of [3] (Parametric).

The supervised univariate approach assumes knowledge of the land cover class of all time series in the training set as well as the initial class prior to change. We fit a univariate Gaussian density for each class at each time step. At test time z-scores are calculated using the appropriate density at each time step assuming the class does not change from its initial assignment.

In the parametric case we fit a harmonic model with a bias, seasonal oscillation component and two higher harmonics over a fixed length look-back window. The model is refit at each time step as the window is shifted. The parametric model is used to make a forecast and the different between observation and prediction is converted to a z-score by dividing by the standard deviation estimated from the model residuals over the window.

Fig. 2 shows a comparison of the histogram and autocorrelation of the z-score series. This was extracted by each method for a single change and no-change time series from the Limpopo data set. Recall that under no change conditions an ideal z-score series should be uncorrelated and Gaussian distributed. The estimated histogram and mean/variance gives an indication of the distribution of the scores. To verify if the scores are independent we can observe the estimated autocorrelation function. A perfectly uncorrelated/independent series should have an autocorrelation that is an impulse centered at zero lag. From inspection of the plots for a no-change time series the autocorrelation plot of the joint estimation methods suggest that the series of z-scores is very close to IID Gaussian. Both the parametric and univariate methods appear to have a much higher degree of correlation between samples in the z-score series. This is expected for the parametric model approach as the small number of harmonics limits the ability of the model to follow fast changes leading to correlated errors. All methods show an increase in correlation in a time series containing land cover change.

B. Change Detection Assessment

To assess the change detection performance of the methods we make use of the Receiver Operator Characteristic (ROC) curve which plots false alarm rate against true positive rate. This allows us to compare methods irrespective of a choice of threshold. Furthermore it does not require us to arbitrarily select an acceptable false alarm rate which may be application specific. We also tabulate the Area Under ROC curve (AUROC) and true alarm rate at a false positive rate of 0.2 in Tables I and II. A ROC curve gives an estimate of the correct classification rates, irrespective of delay, when applying the methods to time series of fixed length and assuming at most one alarm is triggered per series.

All methods requires selection of the slack parameter, k, and both the parametric and the joint Gaussian estimation methods require a finite length look-back window of size W in order to make a forecast. A larger window permits using more historical data at the expense of computational complexity. This is especially true in the joint Gaussian case where a $W \times W$ covariance matrix must be stored and conditioned upon.

To visualize the influence of these parameters Fig. 4 gives the Area Under ROC curve (AUROC) for detection of change/no-change time series as a function of k for several window sizes. For each of the methods it appears that a larger look-back window generally results in improved detection performance. We did not evaluate window size greater than 100 samples as this prevents the detection of changes occurring near the beginning of the series. In a production system this could be increased or even allowed to grow as more

data becomes available. The Gaussian joint density estimation method achieved the highest AUROC on the real data set with a slack parameter setting of around $k = 0.1$. It also appears insensitive to slack parameter selection within the evaluated range. The parametric z-score method was found to work best with a slack of 0 and the supervised method with a slack of 3.0. Recall that the slack parameter is related to the expected change in mean that the z-score series should experience under land cover conversion. These results suggests that the sensitivity of the parametric method to change is already low while the univariate method has a high sensitivity that results in excessive false alarms without sufficient slack.

The ROC curve of was evaluated for a window size of 100 and with the best slack parameter selected from the curves in Fig. 4. Five-fold cross validation was used to estimate the mean ROC curve and its 95% confidence interval. For the Limpopo data set, real change data was used to generate while for the NSW data set synthetic change series were combined with real no-change time series. Fig. 5a gives the curves for the Limpopo data set and Fig. 6a gives the curves for the NSW data set. It can be seen that when considering detection on the Limpopo data set there is no significant difference between each of the assessed methods with the supervised estimation method giving the highest AUROC. On the NSW data set the supervised method did not perform well showing significantly inferior performance to the other methods. This is likely due to the poorly defined land cover classes ('dense' and 'sparse') in this data set. In this case GMM Joint estimation method showed the best performance. Comparison between the Limpopo and NSW data sets illustrates the increase in difficulty in detection within class change compared with between class change.

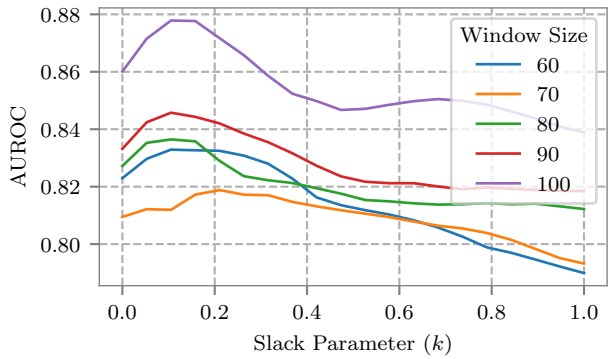
C. Detection Delay Assessment

True/False positive rates give the probability of correct/incorrect detection at any time within the length of the signal. For continuous online detection this does not give any indication on the expected delay until detection. Instead we prefer to compare methods in terms of expected run lengths (i.e. time between alarms). This can be further split into two quantities, the Run Length to False Alarm (RLFA) and the run length to true alarm or Detection Delay (DD). The RLFA is defined as the number of samples after the algorithm starts, either at the beginning of the signal or after reseting due to a previous alarm, to the next alarm. The detection delay is defined as the number of samples between the actual change point and the next alarm. A good detection method strives to maximize the expected RLFA while minimizing the expected DD [10]. These can be considered as the generalization of false alarm rate and true alarm rate to the continuous monitoring case.

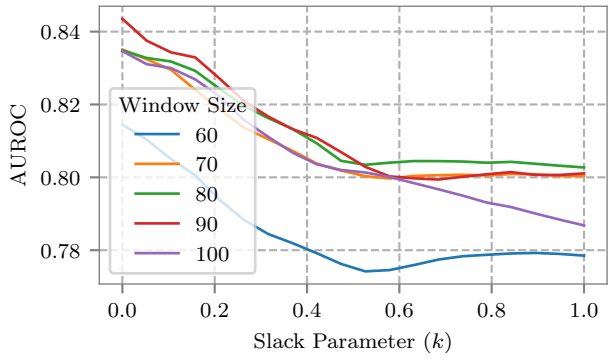
Estimating the distributions of RLFA and DD for a given method using finite length signals presents some challenges. As the z-scores signals are noisy the CUSUM method will always trigger an alarm if it is allowed to run for a sufficiently long time [10]. When estimating run length distributions from finite length signals this duration may be longer than the length of the signal. This can be taken into account by using the concept of censoring from survival analysis. The observation of no alarm gives information that the detection delay was at least as long as the remainder of the signal, known as right censoring. Failure to take this into account, for example by dropping series with no alarm, results in an underestimate of run lengths. To account for the censoring of data we estimate the median detection delay by fitting a Kaplan-Meier estimator [22] a commonly used non-parametric method for estimating a survival function in the presence of censored data.

A survival function was estimated for both the RLFA and DD for each method at a series of thresholds. For RLFA if no alarm was triggered the run length was said to be at least the length of the signal. The RLFA was also considered to be censored when an actual change point occurred. For DD if no alarm was triggered after the change point the run length was said to be at least the difference between the change point and the end of the signal. Fig. 7 compares the Cumulative Distribution Functions (CDF) (1 - survival function) for each method with a single threshold selected to yield a median RLFA of 200. The survival curves give much better insight than the median alone. It can be seen for example that 90% of changes in the Limpopo data set and 70% in the NSW data set are detected with a delay less than 30 samples (120 days) using the GMM joint approach.

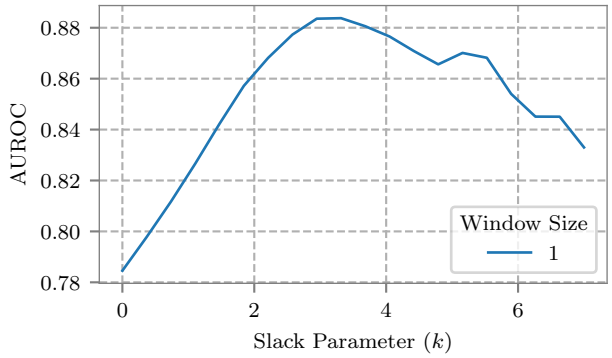
Plotting the median DD vs median RLFA with changing threshold can be considered as a continuous inspection equivalent to a ROC curve. Fig. 5b gives the curves for the Limpopo data set and 6b gives the curves for the NSW data set. Again the mean and 95% confidence intervals for each curve were estimated from five-fold cross validation. From both of these plot it is clear that the GMM joint density estimation methods show a significant advantage in run length when compared with the other methods. Tables I and II give the median detection delays for each method if the threshold is selected to yield a median run length to false alarm 200 samples. For detecting settlement expansion this illustrates that the median detection delay for the joint estimation method is 4 times shorter using the supervised univariate method and around 5 times shorter than the parametric method. With the MODIS sample period of 8 days this corresponds to a median wait of 48 days for the joint method, 232 days for the supervised method and 352 days using the parametric method. This is a significant improvement on supervised CUSUM method of [6] which achieved a detection delay of approximately 400 days applied to synthetic settlement expansion data from the same region. For detecting deforestation in NSW the detection delays were slightly longer in comparison however the joint estimation method still outperformed the parametric method with a median detection delay almost 5 times shorter than the parametric method. The supervised method did not successfully detect enough change points to estimate the median using the Kaplan-Meier method and therefore its curve could not be included. These delays are in the range of those reported in [7] for the detection of deforestation; however, it is difficult to compare as this work does not report the accompanying RLFA.



(a) Joint Gaussian Mixture Model Method



(b) Parametric Method



(c) Supervised Univariate Method

Fig. 4. AUROC as a function of CUSUM slack parameter, k. Plotted for multiple selections of window size for the Limpopo data set, MODIS Band 1.

VI. Conclusion

The forecasting method proposed in paper was shown to give similar performance to two other online methods in terms of detection of a change point within a fixed length signal. The ROC curves of Fig. 5a and 6a suggest that higher false alarm rates must be accepted to yield equivalent true positive rates to other offline methods in the literature applied to the same or similar data sets [17]–[19], [23]. This is a trade-off that must be made for fast, online detection. We show the limitations of the ROC plot for quantifying performance in the continuous inspection case and suggest an alternative of plotting the estimated median run lengths to false alarm and detection delay.

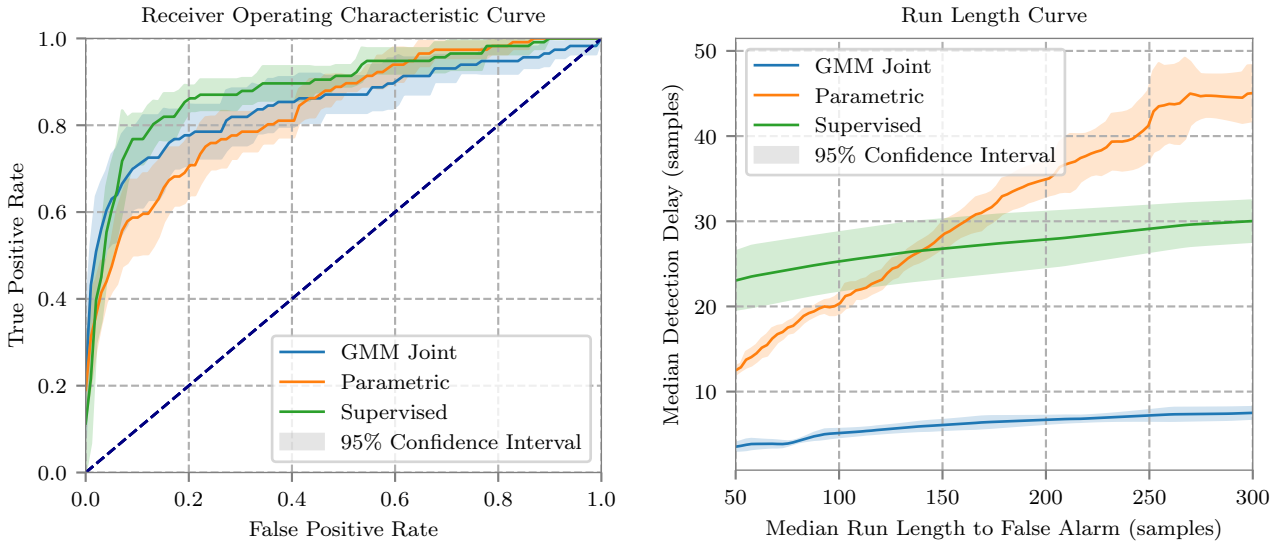
When considering the continuous inspection case the proposed method was shown to exhibit significantly shorter median detection delays for a given median run length to false alarm when compared with the other considered methods. These results suggests the proposed method is a good candidate for land cover change detection system where rapid change detection is important. Future research should consider the generalization of the method to multivariate data to allow multiple bands to be used simultaneously which may be able further reduce the detection delay and occurrence of false alarms.

TABLE I
Results for detection of settlement expansion in Limpopo province using MODIS Band 1

Method	AUROC	TPR at FPR=0.2	DD at RLFA=200
GMM Joint	0.77	0.78 (SD=0.08)	7
Parametric	0.79	0.71 (SD=0.08)	38
Supervised	0.90	0.86 (SD=0.03)	28

TABLE II
Results for detection of vegetation thinning in New South Wales using MODIS Band 1

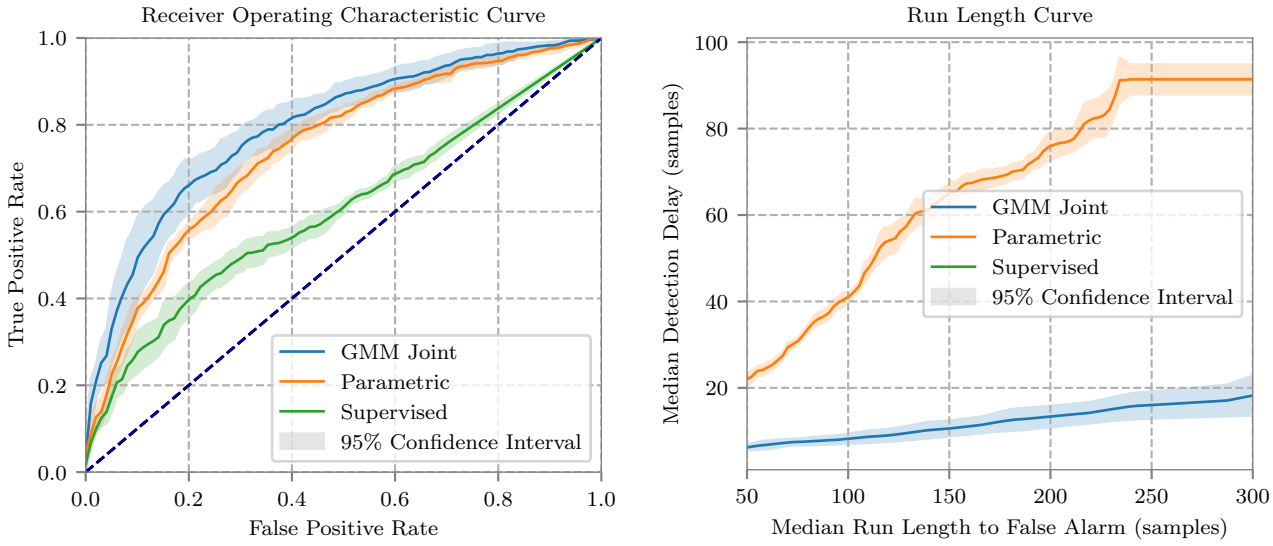
Method	AUROC	TPR at FPR=0.2	DD at RLFA=200
GMM Joint	0.81	0.68 (SD=0.08)	14
Parametric	0.72	0.57 (SD=0.12)	68
Supervised	0.65	0.42 (SD=0.06)	inf



(a) Receiver Operating Characteristic Curves

(b) Detection Delay vs Run Length to False Alarm curves

Fig. 5. ROC curves and Detection Delay curves for each of the considered methods applied to the Limpopo data set using MODIS Band 1. Curves were evaluated with a look-back window size of $W = 100$ and best slack parameter determined from curves. The ROC curves were generated from real change time series and the Detection delay curves from a combination of real no-change and synthetic change time series



(a) Receiver Operating Characteristic Curves

(b) Detection Delay vs Run Length to False Alarm curves

Fig. 6. ROC curves and Detection Delay curves for each of the considered methods applied to the NSW data using MODIS Band 1. Curves were evaluated with a look-back window size of $W = 100$ and best slack parameter determined from curves. Both plots were generating using a combination of real no-change and synthetic change time series.

REFERENCES

[1] R. S. Lunetta and C. D. Elvidge, *Remote sensing change detection*. Taylor & Francis, 1999.

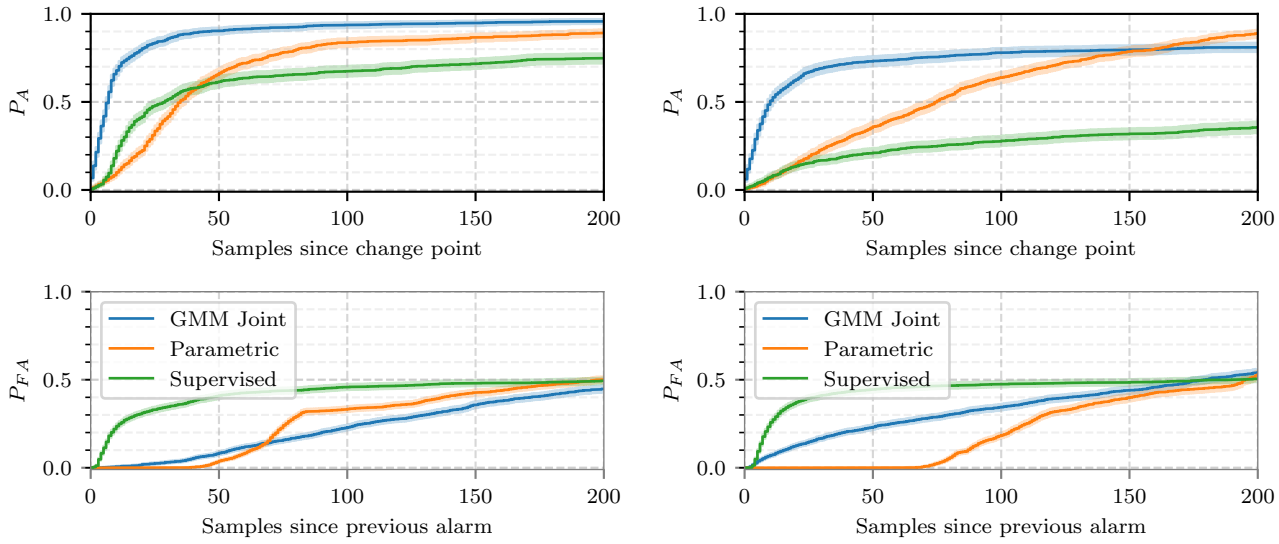
[2] J. Verbesselt, R. Hyndman, A. Zeileis, and D. Culvenor, "Phenological change detection while accounting for abrupt and gradual trends in satellite image time series," *Remote Sens. Environ.*, vol. 114, no. 12, pp. 2970 – 2980, 2010.

[3] J. Verbesselt, A. Zeileis, and M. Herold, "Near real-time disturbance detection using satellite image time series," *Remote Sens. Environ.*, vol. 123, pp. 98–108, 2012.

[4] R. E. Kennedy, Z. Yang, and W. B. Cohen, "Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendrTemporal segmentation algorithms," *Remote Sens. Environ.*, vol. 114, no. 12, pp. 2897–2910, 2010.

[5] V. Chandola and R. R. Vatsavai, "A Gaussian process based online change detection algorithm for monitoring periodic time series," in *Proc. Int. Conf. Data Mining*. SIAM, 2011, pp. 95–106.

[6] T. L. Grobler, E. R. Ackermann, A. J. van Zyl, J. C. Olivier, W. Kleynhans, and B. P. Salmon, "Using page's cumulative sum test on MODIS time series to detect land-cover changes," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 332–336, 2013.



(a) CDFs for Limpopo Data Set

(b) CDFs for NSW Data Set

Fig. 7. Comparison of Cumulative Density Functions (CDFs) for detectors with a fixed thresholds. This gives the probability that an alarm (P_A) or false alarm (P_{FA}) has occurred after a given number of samples. The median delay or run length (plotted in Fig. 5b and Fig. 6b with varying threshold) is the point where the CDF has probability equal to 0.5. Thresholds were selected to yield a median RLFA of 200 samples.

[7] A. Anees and J. Aryal, "Near-real time detection of beetle infestation in pine forests using MODIS data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 9, pp. 3713–3723, 2014.

[8] Z. G. Zhou and P. Tang, "Continuous anomaly detection in satellite image time series based on z-scores of season-trend model residuals," in *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*. IEEE, 2016, pp. 3410–3413.

[9] H. Yin, D. Pflugmacher, R. E. Kennedy, D. Sulla-Menashe, and P. Hostert, "Mapping annual land use and land cover changes using MODIS time series," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3421–3427, 2014.

[10] D. C. Montgomery, *Introduction to statistical quality control*. John Wiley & Sons (New York), 2009.

[11] V. Chandola and R. R. Vatsavai, "A scalable Gaussian process analysis algorithm for biomass monitoring," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 4, no. 4, pp. 430–445, 2011.

[12] P. J. Rousseeuw and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, 1999.

[13] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.

[14] Y. Ritov, "Decision theoretic optimality of the CUSUM procedure," *The Annals of Statistics*, pp. 1464–1469, 1990.

[15] K. C. Steenkamp, K. J. Wessels, S. Archibald, and G. V. Maltitz, "Long-term phenology and variability of southern african vegetation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp*, 2008, pp. 816–819. [Online]. Available: https://doi.org/10.1109/IGARSS.2008.4779474

[16] P. Harrison, "Urbanization: the policies and politics of informal settlement in south africa: a historical perspective," *Africa Insight*, vol. 22, no. 1, pp. 14–22, 1992.

[17] W. Kleynhans, B. P. Salmon, J. C. Olivier, F. van den Bergh, K. J. Wessels, and T. Grobler, "Detecting land cover change using a sliding window temporal autocorrelation approach," in *Proc. IEEE Int. Geosci. Remote Sens. Symp*, July 2012, pp. 6765–6768.

[18] W. Kleynhans, J. C. Olivier, K. J. Wessels, B. P. Salmon, F. Van den Bergh, and K. Steenkamp, "Detecting land cover change using an extended kalman filter on MODIS NDVI time-series data," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 507–511, 2011.

[19] B. P. Salmon, D. Holloway, W. Kleynhans, J. C. Olivier, and K. J. Wessels, "Applying model parameters as a driving force to a deterministic nonlinear system to detect land cover change," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 12, pp. 7165–7176, 2017.

[20] "Carbon farming initiative avoided deforestation 1.1 - methodology determination," Australian Federal Register of Legislation, Mar. 2015.

[21] C. Schaaf and Z. W, "MCD43A4 MODIS/Terra+Aqua BRDF/albedo nadir BRDF adjusted L3 global - 500m v006." in *NASA EOSDIS Land Processes DAAC*, 2015.

[22] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *J. Am. Stat. Assoc.*, vol. 53, no. 282, pp. 457–481, 1958.

[23] B. P. Salmon, J. C. Olivier, W. Kleynhans, and K. J. Wessels, "Using the butterfly effect in a deterministic non-linear system to detect land cover change," in *Proc. IEEE Int. Geosci. Remote Sens. Symp*, July 2014, pp. 4224–4227.

W.C. Olding received the B.Eng. degree in computer systems engineering in 2013 and is currently working toward a Ph.D. degree from the University of Tasmania, Australia.

His research interests include time series analysis, machine learning, signal processing, probabilistic graphical models, image analysis and remote sensing.

J.C. Olivier received the Ph.D. degree in electrical engineering from the University of Pretoria, Pretoria, South Africa, in 1990.

He is currently Professor of Communications Engineering with the School of Engineering and ICT at the University of Tasmania, Australia. He was with Bell Northern Research, Ottawa, ON, Canada; and with Nokia Research Center in the U.S. His research interests are in estimation and detection theory, as well as applications of machine learning.

Dr. Olivier serves as an Editor of the IEEE Transactions on Wireless Communications.

B.P. Salmon received the M.Eng. and Ph.D. degrees in electronic engineering from the University of Pretoria, Pretoria, South Africa, in 2008 and 2012, respectively.

He is currently a Post-doctorate Research Fellow with the School of Engineering and ICT at the University of Tasmania, Australia, and is also affiliated with the Remote Sensing Research Unit, Meraka Institute, CSIR, South Africa. His research interests are information theory, coding theory, and machine learning and graph theory.

Dr. Salmon was the lead author of the paper that won the IEEE Geoscience and Remote Sensing Society 2012 Symposium Prize Paper Award.

W. Kleynhans Dr. Waldo Kleynhans has a Ph.D. in Electronic Engineering from the University of Pretoria and an MBA from Edinburgh Business School, Heriot-Watt University (UK).

His focus is on the application of artificial intelligence and statistical signal (and image) processing in various fields, most notably, satellite remote sensing and telecommunications.

Dr. Kleynhans currently holds adjunct faculty positions with universities and science councils in South Africa as well as the USA.